

GenCore version 5.1.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 18:05:53 ; Search time 1931 Seconds  
(without alignments)  
6118.974 Million cell updates/sec

Title: US-09-829-481-3  
Perfect score: 406  
Sequence: 1 ctctactacaatcactaagt.....aaaaaaaaaaaaaaaaaaaaa 406

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	58	14.3	4900	2	AC115591	AC115591 Dictyoste
C 2	56.4	13.9	1271	9	AF090934	AF090934 Homo sapi
C 3	56.2	13.8	2352	8	ATH304841	ATH304841 Arabidops
C 4	56.2	13.8	156533	2	AC117070	AC117070 Dictyoste
C 5	56	13.8	204557	10	AL669851	AL669851 Mouse DNA
C 6	56	13.8	206345	2	AL844864	AL844864 Mus muscu
C 7	55.8	13.7	797	10	BC010766	BC010766 Mus muscu
C 8	55.8	13.7	3543	9	HSB803534	AL832227 Homo sapi
C 9	55.8	13.7	17294	6	AX345888	AX345888 Sequence
C 10	55.6	13.7	1652	3	AF151644	AF151644 Caenorhab
C 11	55.6	13.7	1717	3	AF121632	AF121632 Drosophi
C 12	55.4	13.6	3024	9	AF146568	AF146568 Homo sapi
C 13	55.4	13.6	27785	2	AC116978	AC116978 Dictyoste
C 14	55	13.5	1432	3	AY069596	AY069596 Drosophi
C 15	55	13.5	1805	3	D91DPP5A	L05617 Dictyosteli
C 16	55	13.5	3152	9	BC034961	BC034961 Homo sapi
C 17	55	13.5	4003	9	BSM804273	AL832962 Homo sapi
C 18	55	13.5	15903	2	AC115612	AC115612 Dictyoste
C 19	55	13.5	160759	2	AC117082	AC117082 Dictyoste
C 20	54.8	13.5	1356	10	BC034886	BC034886 Mus muscu
C 21	54.8	13.5	93491	2	AC116967	AC116967 Dictyoste
C 22	54.8	13.5	209800	2	AC121580	AC121580 Mus muscu
C 23	54.6	13.4	425	6	AX284646	AX284646 Sequence
C 24	54.6	13.4	1451	17	AF118083	AF118083 Homo sapi
C 25	54.6	13.4	2459	3	AY113249	AY113249 Drosophi
C 26	54.6	13.4	113880	3	PFMAL3P4	AL008970 Plasmodi
C 27	54.6	13.4	160759	2	AC117082	AC117082 Dictyoste
C 28	54.6	13.4	171187	2	AC116960	AC116960 Dictyoste
C 29	54.4	13.4	1118	9	BC012595	BC012595 Homo sapi
C 30	54.4	13.4	1577	9	BC011971	BC011971 Homo sapi
C 31	54.4	13.4	1685	9	AK000137	AK000137 Homo sapi
C 32	54.4	13.4	2097	3	AB079606	AB079606 Dictyoste
C 33	54.4	13.4	2789	10	BC026672	BC026672 Mus muscu
C 34	54.4	13.4	3440	9	BC013609	BC013609 Homo sapi
C 35	54.4	13.4	7028	9	BSM803514	AL832207 Homo sapi
C 36	54.2	13.3	785	8	AF257780	AF257780 Populus t
C 37	54.2	13.3	1327	6	AX306545	AX306545 Sequence
C 38	54.2	13.3	1641	9	BC010739	BC010739 Homo sapi
C 39	54	13.3	231807	2	AC131100	AC131100 Mus muscu
C 40	54	13.3	283086	2	AC124723	AC124723 Mus muscu
C 41	53.8	13.3	1588	17	AF116610	AF116610 Homo sapi
C 42	53.8	13.3	1661	9	BC025753	BC025753 Homo sapi
C 43	53.8	13.3	2005	9	HSB803426	AL832119 Homo sapi
C 44	53.8	13.3	6290	6	AX345949	AX345949 Sequence
C 45	53.8	13.3	20622	2	AC115609	AC115609 Dictyoste

ALIGNMENTS

RESULT 1  
AC115591/c  
LOCUS AC115591 4900 bp DNA linear HTG 21-MAR-2002  
DEFINITION Dictyostellum discoideum chromosome 2 map 4713821-4718719 strain  
AX4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AC115591.1 GI:19569972  
VERSION HTG: HTGS\_PHASE2  
KEYWORDS Dictyostellum discoideum.  
SOURCE Dictyostellum discoideum  
ORGANISM Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.  
REFERENCE 1 (bases 1 to 4900)  
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,  
Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,  
Tunggal,B., Cox,E., Quail,M.A., Platzner,M., Rosenthal,A. and

FLI\_CDNA.1 GI:5050233

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Homio sapiens  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
1 (bases 1 to 1271)  
Yu, Y., Zhang, C., Luo, L., Ouyang, S., Zhang, S., Li, W., Wu, J.,  
Zhou, S., Liu, M. and He, F.  
Functional prediction of the coding sequences of 50 new genes  
deduced by analysis of cdna clones from human fetal liver  
Unpublished

```
FEATURES
Source
1. 1271
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HQ0518"
/tissue_type="liver"
/dev_stage="fetus"
```

QY 261 GTTTTATTGTAATTCGTCACGTTTCCAAATTAAAGTCATTTTCGAGGCATACGCAATAATTT 320  
 Db 990 GTATTAAAGTGCACCTCTCTGTGTAATTTCTTAATGAATGACTTTCCAGAAGTGAGATTTGTTA 1049  
 QY 321 TGTAACTTAACAACAGATGCAATAGTTTAAATAAACTTATACCTTTAACTTTAAAAAAA 380  
 Db 1050 TGTCTGGCGCTTTTAAAGGTAAAAATATAAATTAATTTTCACTTAACCTTAATCTAAAAAAA 1109

ACCESSION	AJ304841		i
VERSION	AJ304841.1	GI:11915105	
KEYWORDS	dipi gene; dynamin-like protein Dup1.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		





REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 204557)
TITLE	Howden, P.
JOURNAL	Direct Submission
COMMENT	Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jun 23, 2002 this sequence version replaced gi:20502249. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> RP23-18K13 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a> VECTOR: pBACE3.6
FEATURES	<p>----- Genome Center</p> <p>Center: Wellcome Trust Sanger Institute</p> <p>Center code: SC</p> <p>Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a></p> <p>Contact: humquerry@sanger.ac.uk</p> <p>-----</p> <p>Location/Qualifiers</p> <p>1. .204557</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="X"</p> <p>/clone="RP23-18K13"</p> <p>/clone_lib="RPCI-23"</p> <p>BASE COUNT 62833 a 37986 c 37959 g 65779 t</p> <p>ORIGIN</p> <p>Query Match 13.88; Score 56; DB 10; Length 204557;</p> <p>Best Local Similarity 55.78; Pred. No. 0.34; Indels 0; Gaps 0;</p> <p>Matches 107; Conservative 0; Mismatches 85;</p> <p>QY 215 GTTGTCTAGTGAATCGATTATTATTCGCATATGAGACCCGTTTTTATTGTAATA 274</p> <p>                                                     </p> <p>Db 161602 GTTGCTCTTCAGGACATCGAAGATTGTGTATCTTGGCCAGCTCTCTGCAATATA 161543</p> <p>                                                     </p> <p>QY 275 TCGTCAGTTTCCAATAAAGTCATTTCGAGCCATCTGAATAATTGTTGTAATCAACAC 334</p> <p>                                                     </p> <p>Db 161542 GCATCTCTACCATTTAGATCATGTGAACCAAGGGGAAACATCTGTGACAGACATCT 161483</p> <p>                                                     </p> <p>QY 335 AGATCAATAGTTTAAATAAATTATCTTAACTTTTTAAAAAATAAAAAAAAAAAAA 394</p> <p>                                                     </p> <p>Db 161482 TGAATAACAGAGCAATAAAGTCTGCCTAGATCTAGACAGATAAAAAAAAAAAAA 161423</p> <p>                                                     </p> <p>QY 395 AAAAAAAAAAAAA 406</p> <p>                                                     </p> <p>Db 161422 AAAAAAAAAAAAA 161411</p> <p>                                                     </p> <p>RESULT 6</p> <p>AL844864/c 206345 bp DNA linear HTG 29-JUL-2002</p> <p>LOCUS Mus musculus chromosome X clone RP23-360L8, *** SEQUENCING IN</p> <p>DEFINITION PROGRESS ***; 15 unordered pieces.</p> <p>ACCESSION AL844864</p>
VERSION	AL844864.2 GI:22022618
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Mus musculus
ORGANISM	Mus musculus
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 206345)
TITLE	McLay, K.
JOURNAL	Direct Submission
COMMENT	Submitted (28-JUL-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 30, 2002 this sequence version replaced gi:22003281.
FEATURES	<p>----- Genome Center</p> <p>Center: Wellcome Trust Sanger Institute</p> <p>Center code: SC</p> <p>Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a></p> <p>Contact: humquerry@sanger.ac.uk</p> <p>-----</p> <p>Location/Qualifiers</p> <p>1. .206345</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="X"</p> <p>-----</p> <p>Location/Qualifiers</p> <p>1. .206345</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="X"</p>

RESULT 8  
HSM803534

RESULT 7  
BC010766

```
LOCUS      HSM803534      3543 bp      mRNA      linear      PRI 10-JUL-2002
DEFINITION Homo sapiens mRNA; cDNA DKF2p686p1536 (from clone DKF2p686p1536).
ACCESSION  AL832227
VERSION    AL832227.1  GI:21732773
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3543)
AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C. and Wiemann,S.
TITLE     Direct Submission
JOURNAL   Submitted (09-JUL-2002) I, D-85764 Neuberberg, GERMANY
COMMENT   Clone from S. Wiemann. Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKF2p686p1536) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES   Location/Qualifiers
            source
            1..3543
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKF2p686p1536"
            /tissue_type="cDNA-collection"
            /clone_lib="586 (synonym: hlcc3). Vector pSport1_Sfi; host
            DH10B; sites SfiI + SfiIb"
            /dev_stage="adult"
            polyA_site 1077 a 656 c 623 g 1187 t
BASE COUNT 1077 a 656 c 623 g 1187 t
ORIGIN
Query Match 13.7%; Score 55.8; DB 9; Length 3543;
Best Local Similarity 57.0%; Pred. No. 0.6;
Matches 102; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 228 GAATATCGGATTATTTGCCATATGAGACCCGTTTATTAATGAATATCGTCAGTTTCCA 287
DB 3316 GATAATCGGTGTGTTATTTATTTGCGAGAAATGATGATATATCATACAAAGCACACA 3375
QY 288 ATTAAGATCATTTCCAGGCATACACTGAATAATTTGTGAATCTAACACAGATGCAATAGTT 347
DB 3376 ATATATATATTTCTTTCGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3435
QY 348 TAAATAAACTTATACCTTAACTTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATA 406
DB 3436 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 3494
RESULT 9
AX345888/c 17294 bp DNA linear PAT 01-FEB-2002
LOCUS      AX345888      17294 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION Sequence 959 from Patent WO0200928.
ACCESSION  AX345888
VERSION    AX345888.1  GI:18493774
KEYWORDS   synthetic construct.
SOURCE     synthetic construct.
ORGANISM   artificial sequences.
REFERENCE  1
AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
TITLE     Diagnosis of diseases associated with the immune system
JOURNAL   Patent: WO 0200928-A 959 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES   Location/Qualifiers
            source
            1..17294
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="Chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 4922 a 203 c 3691 g 8478 t
```

```
ORIGIN
Query Match 13.7%; Score 55.8; DB 6; Length 17294;
Best Local Similarity 53.4%; Pred. No. 0.5;
Matches 117; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 184 CTTTTCGAGAGAACCTTCAGGACACCTGCGTTTCTATAGTGAATAATCCGATTATT 243
DB 280 CTTTAACCTATAAAAAAATAAACCCCTATTAAATTTAAATCACCATCAATTCGATT 221
QY 244 TGCATAATGGAGACCCGTTTATTGAATATCGTCAGTTTCCATTAAGATCAATTCGA 303
DB 220 TTCTATTAATCGTACAAATTTTTTTTTTTTTTTTTTTTTTTTACTCAATCAACAACA 161
QY 304 GCCATCTGAATAATTTGTGAATCTTACACACAGATGCAATAGTTTAAATAAACTTATACT 363
DB 160 ACTATAAATTAATTTTCAAAAAATTAATAATATAACCATATAAAAAAATAAATAAC 101
QY 364 TAACTTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 402
DB 100 TAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 62
RESULT 10
AF151644      1652 bp      mRNA      linear      INV 27-JUL-2001
LOCUS      AF151644      1652 bp      mRNA, complete cds.
DEFINITION Caenorhabditis elegans ionotropic GABA receptor subunit UNC-49C
ACCESSION  AF151644
VERSION    AF151644.1  GI:5354173
KEYWORDS   Caenorhabditis elegans.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 1652)
AUTHORS   Hamber,B.A., Beg,A.A., Twyman,R.E. and Jorgensen,E.M.
TITLE     The Caenorhabditis elegans unc-49 locus encodes multiple subunits
            of a heteromultimeric GABA receptor
JOURNAL   J. Neurosci. 19 (13), 5348-5359 (1999)
MEDLINE   9307327
PUBMED    10377345
REFERENCE  2 (bases 1 to 1652)
AUTHORS   Hamber,B.A. and Jorgensen,E.M.
TITLE     Direct Submission
JOURNAL   Submitted (17-MAY-1999) Biology, University of Utah, 257S 1400E,
            Salt Lake City, UT 84112-0840, USA
FEATURES   Location/Qualifiers
            source
            1..1652
            /organism="Caenorhabditis elegans"
            /strain="Bristol N2"
            /db_xref="taxon:6239"
            /chromosome="III"
            /map="between lin-19 and mel-23"
            1..1652
            /gene="unc-49c"
            114..1460
            /gene="unc-49c"
            /codon_start=1
            /product="ionotropic GABA receptor subunit UNC-49C"
            /protein_id="AAD42386.1"
            /db_xref="GI:5354174"
            /translation="MARPTLIIVLSAHLCHVVVQDEDSHINTOLLSSVLDRLTNR
            TTYDKRLRPYRGKPDVGTIVHSSTISAVSEVDMFTLDFYMRQWOPRLAFGLSD
            LGLSKELDSLTVGDVLDRLWKPDFTFPNEKKKSFHFLATTHNSFLRIEGDGVYTSOR
            LTVATCPMDLKLFPMDSOHCKLEISYASTAEIEYKWTCKEPCNSTAVKADANIE
            LSSYKFTKICQKRTLASTSGTYSRLRVSFIFRDSGFYFLQIFFPASLVVLWSLSE
            WINRDSAPSRTLGTMVLTETHLMTGNRLPPVAIVRAVDVFLGFCVLLVLALIE
            YACVAYSKKKNEDRRREKHEKPPPTPDILHDLAECTCAACTCAAPTSTIAIVIKOSN
            RFCVSHSHDIVSRAAFPFLVFLIFLFWLILLYLKYKRLPYISEHGEGRCDAPDLH"
BASE COUNT 456 a 413 c 323 g 460 t
ORIGIN
```

Query Match	13.7%;	Score 55.6;	DB 3;	Length 1652;	
Best Local Similarity	55.1%;	Pred. NO. 0.72;			
Matches 109;	Conservative	0;	Mismatches 89;	Indels 0;	Gaps 0;
QY	209	ACCTGGCTTGGCTATAGGTGAAATCCGATTTATTTGCCATATGAGACCGCTTTTAT	268		
Db	1450	ACCTTCATTAATCTCAATCCAACTTCCTCATCTATTTCCATTCGAATATCTCTTTTCT	1509		
QY	269	TGAATATCGTCAGTTCCCAATTAAGCATCTTTGCGACCATACTGAATAATTTGTAATCT	328		
Db	1510	TGCACAGAGACCCCTTTTTCGTTTTTTTATTGATTATTTATTACGGATTTTATAGATAAT	1569		
QY	329	AACAACAGATGCATATAGTTTAATAAATCTATATCTTTAACTTTTAAAAA	388		
Db	1570	GCACAGATGCCTCATGTCCTCAATAATTTATTTAAATTTGTCGAAAAA	1629		
QY	389	AAAAA	406		
Db	1630	AAAAA	1647		
RESULT 11					
AY121632					
LOCUS	Drosophila melanogaster	GM01935	full insert cDNA.		
DEFINITION	AY121632				
ACCESSION	AY121632.1	GI:21464311			
VERSION	FLI-CDNA.				
KEYWORDS	fruit fly.				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1717)				
AUTHORS	Stapleton,M., Brokstein,P., Hong,L., Agbavani,A., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G., Miranda,A., Mungall,C.J., Nunoo,J., Pachleb,J., Paragas,V., Park,S., Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M. and Celisner,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-JUN-2002) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA				
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory Berkeley, CA 94720				
	This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site ( <a href="http://fruitfly.berkeley.edu">http://fruitfly.berkeley.edu</a> ) or send email to <a href="mailto:cdna@fruitfly.berkeley.edu">cdna@fruitfly.berkeley.edu</a> .				
FEATURES	source	1..1717			
	Location/Qualifiers				
	1..1717				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	<1..1717				
	/gene="Rp11"				
	/db_xref="FLYBASE:FBgn0019938"				
	<1..1629				
	/gene="Rp11"				
	/note="Longest ORF"				
gene					
CDS					
Query Match	13.7%;	Score 55.6;	DB 3;	Length 1717;	
Best Local Similarity	82.1%;	Pred. NO. 0.72;			
Matches 64;	Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0;
QY	329	AACAACATGCAATAGTTTAAATAAATCTTATATCTTAACTTTTAAAAA	388		
Db	1638	AATATAATTGAAATTTTAAATAAATCAATACATTAACGTTAAAAA	1697		
QY	389	AAAAA	406		
Db	1698	AAAAA	1715		
RESULT 12					
AF146568					
LOCUS	AF146568				
DEFINITION	Homo sapiens MLL1 protein mRNA, complete cds; nuclear gene for mitochondrial product.				
ACCESSION	AF146568				
VERSION	AF146568.1	GI:6103438			
KEYWORDS	Homo sapiens.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3024)				
AUTHORS	Zemskova,M.Y., Lilly,M. and Escher,A.P.				
TITLE	MLL1, a novel human gene encoding mitochondria located protein promoting cell survival				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3024)				
AUTHORS	Zemskova,M.Y. and Escher,A.P.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-APR-1999) Center for Molecular Biology and Gene Therapy, Loma Linda University, 11085 Campus Street, Loma Linda, CA 92350, USA				
FEATURES	source	1..3024			
	Location/Qualifiers				
	1..3024				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_line="glioma LNZ308"				
	1..71				
	72..1232				
	/function="involved in promoting cell survival"				
	/codon_start=1				
	/product="MLL1 protein"				
	/protein_id="AAFO3602.1"				
	/db_xref="GI:6103439"				
	/translation="MEDCLAHGEKVSOELKPLHKLQMLLSOPVTVOAFRECTLET				
	TVHASSWNKILVPLVLLRQMLLETRLGOELPSALLQFGVTYLEDYSAEYLIQGGWG				
	TVFSLESEEEYPGITAEDSNDIYLLPSDNGOVSPSPSTVTSWQSESLPVSLSA				
	QSWHTESLPVSGIPESWQCIAMDPEEVLSNGAGEKSENNSNDIVHVEKEEVP				
	GMEAAVAVSVLPARELOALPEAPALPHITATSLTGTPDTPTVITVHKKSPATS				
	LFVELDEEVKAATTEPTEVEVPALEPTLTISEKINAREESLVLELSPASEKKP				
	VPPSECKSRLSPAGEM				



```
3'UTR      1233..3024
BASE COUNT 850 a 701 c 728 g 745 t
ORIGIN

Query Match      13.6%; Score 55.4; DB 9; Length 3024;
Best Local Similarity 61.4%; Pred. No. 0.73;
Matches 89; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 262 TTTTATTGAATATCGTCAGTTTCAATTAAGTCATTTCCGAGCCATACCTGAATAATTTT 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2859 TTTCTTTTGAATTTATGTTTCAATAAATCCCTAAAGCAATATTTAAATATGG 2918

QY 322 GTAATCTAACAGATGCAATAGTTTAAATAAATCTATATCTAATCTTTTAAAAA 381
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2919 TCAAAAAA 2978

QY 382 AAAAAAAAAAAAAAAAAAAAAA 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2979 AAAAAAAAAAAAAAAAAAAAAA 3003

RESULT 13
AC116978
LOCUS      Dictyostelium discoideum chromosome 2 map 4846697-4874480 strain
DEFINITION AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION  AC116978
VERSION     AC116978.1 GI:20042933
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE      Dictyostelium discoideum.
ORGANISM    Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1 (bases 1 to 27785)
AUTHORS     Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
            Sequence and Analysis of Chromosome 2 of Dictyostelium
            Unpublished
            The Dictyostelium Genome Sequencing Consortium
            2 (bases 1 to 27785)
            Baumgart,C.
            Direct Submission
            Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
            3 (bases 1 to 27785)
            Baumgart,C.
            Submitted (31-MAY-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
            CDS predictions from GenBank may contain errors. Further Information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the University Cologne, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml)
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
            * Note: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            Location/Qualifiers
            1..27785
            /organism="Dictyostelium discoideum"
            /strain="AX4"
            /db_xref="taxon:44699"
            /chromosome="2"
            /map="4846697-4874480"
            678..1491
            /note="ORF_ID:dd_01248"
            /pseudo
            CDS
            join(2445..2467,2559..3372)

FEATURES
            source
            1..27785
            /product="hypothetical protein"
            /protein_id="AA045261.1"
            /note="ORF_ID:dd_01251"
            /pseudo
            join(6336..6630,6711..11386,11471..14377)
            CDS
            /note="ORF_ID:dd_00705"
            /codon_start=1
            /product="Hypothetical 178.4 kDa protein"
            /protein_id="AA045260.1"
            /db_xref="GI:21281373"
            /translation="MMTAAAKEFLFKICGEYPMYLEKDFKIAFKKGTINLNNLEL
            NIKKINDLYGVSPMLASGYIGNLTMTIPYTSLLDTPSEFNISLELTFVSSNDYEK
            YVLEGLKQLAYLDHSSTKNDNNNNNNNNVNYVYDEDEIDETETKQGGIOGLE
            ALTEIVKLLDKVSARTDNCITITVEYDOSRNKSIITIVELIESIYRASSTGPPNYF
            NNNNNNNNNNNNTSSQEKIDYFLKSTVLKNSHLLITDKYTGNTILLEEFYKNPENI
            IPKSKFDLIESKVDSSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            KIINTNTNNNELPFLVDFCFQWTIVLSPHLSIIDSFSNFKPQKSTTTTTT
            TTTTNTSIPNITKRQSIKSNQNLNKLWQSIKIYHKDRKENVYNEEIPMWI
            DKLNPQPKRDYLALEFDINTSLKQNNIKKMLTIKLLITNYVNSTCEGLIN
            EITKPISSNSPSSSSSIVSNLSNKLKLNSTNNNNNNNNNNNNNNNNNNNNNNNN
            QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
            PELLKLSFTSEMKEVEKEEKLKLYLNRIEKNLIKNNLLKQKQDLRLOR
            KELOKQLOHQHQHOSDYYPREYDYVDSDEGDEYVYHEGDSKNNQNNNNNN
            NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            STFEIQKLLTQLEISFTQSQSVIRKFPKFPFEPENISLVLDLKKKFK
            GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            LLNNKNTGNSDKVDQHQDEKDEVOEQEDKIDINDYSGGSGFNIITNDVS
            DDIDDDIDFEMTDDIDYEDGNVKIPONIPOLPSPPTPTPIPOQOQOQOPO
            PPQOQOQOQOHOQNGTSPNFHNPFSKSPKHKGPSTFVSTSEGVDSWDEAPE
            OETINPHSSIESSRYAIIIFLOKLKTSQYQSLQSLQSLQSLQSLQSLQSLQSLQ
            VSSFINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            ILPKRNGSPQVKSFNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            FI IQEDAPKSSFNSSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
            VYGRDISFVSSVSKVLNTLTKDNKPLPIFKTIISQERNSTDMTISVTFNRD
            TTGFLDNIHGNVYFVKGTIYHKLTQWYSGRIENFNDSTNNNNNNNNNNNNNNNN
            NNNNNNTTSNENPKDEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE
            INFVDSISVTPSDNLSAAVLFSDVKLKIPTSFGSCIGONSYLYVMDLQFLNLN
            XPSKSSSSNNLNNKLYGVVHLDIETDITSSNDNDHYPSLKIDNNHLLSIN
            ACSDFYILNQLNLFGLDPLAEVADLLITQIGSELNKLKESIDILINETFFQTA
            EOLQOQOQOQOQSSNNNLSLSSPGGPTINSNNNNNNNNNNNNNNNNNNNNNNNN
            FEEVLRGDNDFDDSSSVYIYSPLDMSNLSKLSSETSCNENOMINEIHTDI
            SSHTGFSNLPIMGSIIMIEDYDHDKLETYPSTTQWPHROFYNDIDPPSYLQLOQP
            TOSSSYQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
            PSDEPLPSHPKPSAKSMKFRKWLSTIKFGLDWERESTILEPSSPSLAGKKEKQ
            SQOQOQPHKSGRQONCFVEFILLSNVIIRDREDDNELYAKRMSLHISDVTMDHPT
            SLMNKFLCSDTNTRYQOQPMIKLIVETVARDIARPLQELNKLKILILIRNIDODT
            VSEMVNYFSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
            TTSITNTTTTSSISPSIIISPLSTQSTFTTGGSSSSSSSSSSSSSSSSSSSSSS
            SSINTSSNVTPIPKKSSSSSSSSSQTISRSIPEDKAOTTKQSEQIYFQSVETLP
            IRLKVDYKPKKVVDYHSLTTCGVAELLNLPLEGAIFNLAKTLKNGVGGSAFLPQIAK
            IWPPIIOTOLLGYVSGVGLNSLHIGEGLANLILPYEOKKDKGNTLKGKRTT
            FLKNIIVETLSMGAKAVAGTOGLETADNVLSPKQOHSQHSQHSQHSQHSQHSQ
            SSTOKHSSSSSHGRFNHQSILHSFNSPTQOLSPASISLHLLHLLHLLHLLHLLH
            DTREGIQHAYESVRELKSAATIIAIPMKREYHQKGTGKGVKSIQVAVPITAIRPMIG
            ITCGVSKTLGVNRQIDPNKKKLEMDNRYK"
            complement(join(14916..15729,15822..15844))
            /note="ORF_ID:dd_00706"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AA045261.1"
            /db_xref="GI:21281374"
```



Job time : 2057 secs

```
CDS
69..878
/gene="CG1643"
/note="Longest ORF"
/codon_start=1
/product="LD34980p"
/protein_id="AAL39741.1"
/db_xref="GI:17862528"
/db_xref="FLYBASE:FBgn0029943"
/translacion="MAHDREVLRIWEGQIGICFQADREIVGKPEPEPYLMISRLSY
LPIVTKVRKYFYRTSAEHQDGVWDFNGTFLRHYPGVLYDLHHPEDSTPWCL
TIHFSFPEDMLVKLSHESYKLEADVLKHLGLVISAMOKKDHNLQWLGLV
NEKFDQFWAVNRRLMEPYGDLESFKNIPLRITYDDDTYTKLISPSYVGOKKSLAD
LMAELSTPVRRAVGCRTHGIDLHEETQLOWMSEHLSYPDNFLHLSVDYKDV"

BASE COUNT      441 a   312 c   301 g   378 t
ORIGIN
```

```
Query Match      13.5%; Score 55; DB 3; Length 1432;
Best Local Similarity 66.4%; Pred. No. 0.95;
Matches 79; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 288 ATTAAGTCATTTCAGCCATCACTCAATAATTTGTATCTAACAACAGATGCAATAGTT 347
DB 1302 ACTAAATGACATCACGCCACCACTAATGATTAAATGGAAACTATTTTATATCGACCCCAATC 1361

QY 348 TAAATAAACTTATCTAACTTTTAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
DB 1362 RATATATATATACGAAAGTTATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1420
```

```
RESULT 15
DDIDPP5A      1805 bp      DNA      linear      INV 11-MAR-1994
LOCUS      Dictyostelium purpureum (Dpp5) DNA sequence, repeat region.
DEFINITION
ACCESSION      L05617
VERSION      L05617.1 GI:167774
KEYWORDS      repeat region.
SOURCE      Dictyostelium purpureum (strain DPA) DNA.
ORGANISM      Dictyostelium purpureum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE      1 (bases 1 to 1805)
AUTHORS      Kiyosawa,H., Hughes,J.E., Podgorski,G.J. and Welker,D.L.
TITLE      Small circular plasmids of the eukaryote Dictyostelium purpureum
define two novel plasmid families
JOURNAL      Plasmid 30 (2), 106-118 (1993)
MEDLINE      94052367
PUBMED      8234483
FEATURES
source      Location/Qualifiers
1..1805
/organism="Dictyostelium purpureum"
/strain="DPA"
/db_xref="taxon:5786"
```

```
repeat_region 320..477
BASE COUNT      783 a   159 c   202 g   661 t
ORIGIN
```

```
Query Match      13.5%; Score 55; DB 3; Length 1805;
Best Local Similarity 61.5%; Pred. No. 0.92;
Matches 88; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 264 TTATTGAATATCGTCAGTTCCCAATTAAGTCATTTCGAGCCATCACTCAATAATTTGT 323
DB 1154 TTAATTCAATTTTGATATATAAAAAATAAAAAATAGTATATTGATAAACTATAAATTTA 1213

QY 324 AATCTAACACAGATGCAATAGTTTAAATAAACTTATACTTAACTTTTAAAAAIAAAAA 383
DB 1214 AAAAAAAAAAAAAAAAAAATTTTAAAAAIAAAAAATATATATGCGGATATATAAATAAAAA 1273

QY 384 AAAAAAAAAAAAAAAAAAAAAA 406
DB 1274 AAAAAAAAAATAAAAAAAAAAA 1296
```



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 18:00:58 ; Search time 230 Seconds  
(without alignments)  
3975.267 Million cell updates/sec

Title: US-09-829-481-3  
Perfect score: 406  
Sequence: 1 ctctactacaatcactaagt.....aaaaaaaaaaaaaaaaaaaa 406

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID52/qcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID52/qcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID52/qcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID52/qcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID52/qcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID52/qcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID52/qcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID52/qcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID52/qcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID52/qcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID52/qcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID52/qcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID52/qcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID52/qcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID52/qcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID52/qcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID52/qcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID52/qcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID52/qcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID52/qcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID52/qcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID52/qcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID52/qcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID52/qcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	406	100.0	406	24	AAH77209		Arthropod defensin
2	76	18.7	386	24	AAH77210		Arthropod defensin
3	59.2	14.6	230	23	ABV57465		Human prostate exp
4	59.2	14.6	236	23	ABV56953		Human prostate exp
5	58	14.3	574	23	ABV57502		Human prostate exp
6	57	14.0	497	23	ABV58273		Human prostate exp
7	56	13.8	381	23	ABV58662		Human prostate exp
8	55.8	13.7	17294	24	ABL32986		Human immune syste
9	55.2	13.6	325	22	AAS29049		CDNA encoding for

10	55.2	13.6	621	22	AAS29124		CDNA encoding for
11	54.6	13.4	351	24	AAH77211		Arthropod defensin
12	54.6	13.4	425	22	AAS60450		Human cancer agent
13	54.6	13.4	2736	24	ABO54618		Human ovarian anti
14	54.4	13.4	372	23	ABV37528		Human prostate exp
15	54.4	13.4	447	23	ABV10351		Human prostate exp
16	54.2	13.3	266	23	ABV07596		Human prostate exp
17	54.2	13.3	1327	24	AD24775		Glycine max ankryl
18	53.8	13.3	309	23	ABV44994		Human prostate exp
19	53.8	13.3	431	23	ABV04399		Human prostate exp
20	53.8	13.3	461	24	AAH77208		Arthropod defensin
21	53.8	13.3	6290	24	ABL33047		Human immune syste
22	53.6	13.2	154	22	AAL23561		Human breast cance
23	53.6	13.2	193	22	AAL14699		Human breast cance
24	53.6	13.2	1812	21	AA16649		Human secreted pro
25	53.4	13.2	6189	22	AAS46600		Tumour suppressor
26	53.4	13.2	6189	24	ABL34593		Human metastasis a
27	53.2	13.1	300	22	AAH70049		Human cervical can
28	53	13.1	959	24	ABK63920		CDNA encoding huma
29	52.8	13.0	3581	15	AAQ44391		Sequence of murine
30	52.6	13.0	3293	22	AAS25944		Human CDNA encodi
31	52.6	13.0	5198	24	ABL32279		Human immune syste
32	52.4	12.9	165	22	AA184674		Human polynucleoti
33	52.4	12.9	958	22	AAC92233		Human polynucleoti
34	52.4	12.9	958	24	AAD22188		Apoaequorin-encodi
35	52.4	12.9	959	9	AAH81534		Aequorea victoria
36	52.4	12.9	1994	22	AAH48294		pAQ440 aequorin ge
37	52.4	12.9	2175	21	AAC79978		Honeybee alpha-glu
38	52.2	12.9	368	22	AAS60046		Human secreted pro
39	52.2	12.9	426	22	AA188844		Human cancer agent
40	52.2	12.9	441	22	AA191749		Human polynucleoti
41	52.2	12.9	504	22	AA185190		Human polynucleoti
42	52.2	12.9	612	22	AAH71471		Human cervical can
43	52.2	12.9	815	21	AAC79968		Human secreted pro
44	52.2	12.9	1387	22	AAS29132		CDNA encoding for
45	52.2	12.9	3957	22	ABA06411		Human CDNA SEQ ID

# ALIGNMENTS

RESULT 1  
AAH77209  
ID AAH77209 standard; cdna; 406 BP.  
XX  
AC AAH77209;  
XX  
DT 30-JAN-2002 (first entry)  
XX  
DE Arthropod defensin cdna from Vaejovis carolinianus clone ikslc.pk0.  
XX  
KW Arthropod: defensin; ikslc.pk0; antimicrobial; microbial pathogen;  
KW gene therapy; ss.  
XX  
OS Vaejovis carolinianus.  
XX  
PH Key Location/Qualifiers  
FT CDS 44..229  
FT /\*tag= a  
FT /product= "Defensin"  
XX  
PN EPI146052-A2.  
XX  
PD 17-OCT-2001.  
XX  
PF 12-APR-2001; 2001EP-0303488.  
XX  
PR 14-APR-2000; 2000US-197279P.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
PI Presnail JK, Weng Z, Wong JF;  
XX

```

DR WPI; 2002-001056/01.
XX P-PSDB; AAG77932.
PT New polynucleotides encoding Arthropod defensin polypeptides for
PT improving the microbial resistance of plants and animals by gene
PT therapy -
XX
XX Claim 6; Page 18; 22pp; English.
XX
XX The sequence represents a novel Arthropod defensin from V. carolinianus
CC clone ikslc.pk0. The invention relates to novel polynucleotides encoding
CC Arthropod defensin polypeptides. The polypeptides and polynucleotides of
CC the invention have antimicrobial activity. The polynucleotides may be
CC used to genetically engineer cells and organisms, especially plants, to
CC alter their resistance to a wide range of microbial pathogens. The
CC polynucleotides may also have a use in gene therapy.
XX
XX Sequence 406 BP; 143 A; 66 C; 67 G; 130 T; 0 other;
SQ
Query Match 100.0%; Score 406; DB 24; Length 406;
Best Local Similarity 100.0%; Pred. No. 3.le-64;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCTACTACAATCACTAAGTCTCTTCTCCACTCAGCTTCAAGATGAATCCATAGCTAT 60
DB 1 CTCTACTACAATCACTAAGTCTCTTCTCCACTCAGCTTCAAGATGAATCCATAGCTAT 60
QY 61 TATTTTCATCGTCTTGTTGCTCTCTGTATTTTGGAGGATGGGATTTGAGAAGCTGGTTT 120
DB 61 TATTTTCATCGTCTTGTTGCTCTCTGTATTTTGGAGGATGGGATTTGAGAAGCTGGTTT 120
QY 121 TGGATGTCCTTTAATGCAGGAAATGCCATAGACATTCGAAAGTATTCGCTGAGAG 180
DB 121 TGGATGTCCTTTAATGCAGGAAATGCCATAGACATTCGAAAGTATTCGCTGAGAG 180
QY 181 AGGCTTTTGCAGAGAACTTTTCAGGACAACTGCGTTTGCTATAGGTGAAATCCGATTT 240
DB 181 AGGCTTTTGCAGAGAACTTTTCAGGACAACTGCGTTTGCTATAGGTGAAATCCGATTT 240
QY 241 ATTTGCCATATGAGACCCGTTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTT 300
DB 241 ATTTGCCATATGAGACCCGTTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTT 300
QY 301 CGAGCCATCTGAATAATTTTGTAATCTTAACAACAGATGCAATAGTTTAAATAAATAT 360
DB 301 CGAGCCATCTGAATAATTTTGTAATCTTAACAACAGATGCAATAGTTTAAATAAATAT 360
QY 361 ACTTAACCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406
DB 361 ACTTAACCTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406
RESULT 2
AAH77210
ID AAH77210 standard; cDNA; 386 BP.
XX
XX AAH77210;
XX
XX 30-JAN-2002 (first entry)
XX
XX Arthropod defensin cDNA from Argiope sp. clone aotlc.pk0 #1.
XX
XX Arthropod; defensin; aotlc.pk0; antimicrobial; microbial pathogen;
XX gene therapy; ss.
XX
XX Argiope sp.
XX
XX Key Location/Qualifiers
XX CDS 14..196
XX /*tag= a
XX /product= "Defensin"
XX
XX EP1146052-A2.

```

```

XX 17-OCT-2001.
XX
XX 12-APR-2001; 2001EP-0303488.
XX
XX 14-APR-2000; 2000US-197279P.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Presnail JK, Weng Z, Wong JF;
XX
XX WPI; 2002-001056/01.
XX P-PSDB; AAG77933.
XX
XX New polynucleotides encoding Arthropod defensin polypeptides for
XX improving the microbial resistance of plants and animals by gene
XX therapy -
XX
XX Claim 6; Page 18; 22pp; English.
XX
XX The sequence represents a novel Arthropod defensin from Ariope sp.
CC clone aotlc.pk0. The invention relates to novel polynucleotides encoding
CC Arthropod defensin polypeptides. The polypeptides and polynucleotides of
CC the invention have antimicrobial activity. The polynucleotides may be
CC used to genetically engineer cells and organisms, especially plants, to
CC alter their resistance to a wide range of microbial pathogens. The
CC polynucleotides may also have a use in gene therapy.
XX
XX Sequence 386 BP; 140 A; 81 C; 67 G; 94 T; 4 other;
SQ
Query Match 18.7%; Score 76; DB 24; Length 386;
Best Local Similarity 54.4%; Pred. No. 3.8e-05;
Matches 178; Conservative 0; Mismatches 141; Indels 8; Gaps 1;
QY 78 TTGCTCTCTGTATTTGGAGGATGGGATTTGAGAGCTGTGTTTGGATGTCCTTTAATG 137
DB 45 TAGTCGTCTGTGCTTTTGGCCACAGTGGCGCTGGAAGCTGTTTCGGCTGCCCTTCGACC 104
QY 138 CAGGAAATGCCATAGACATTCGAAAAGTATTCGTCGTAGAGGAGGCTTTTCGACAGGAA 197
DB 105 AGATGCGATGTGCACATCATTTGCAGGAGCATCAAAATACAGGGGAGGATCTGCACCACT 164
QY 198 CTTTCAGCACAACTGCGTTTGTCTATAGTCAAAATCCGATTTATTTGCCATATGAGA 257
DB 165 TATTCAGCGCACCTGCAAGTGTACGGATGATGACCCCTCCCTCTCAGACAGGAGC 224
QY 258 CCCCTTTTATTTGAATATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCATCTGAATAA 317
DB 225 CCACACCTT-----TCATTGACATCCGATTTCCCAATTCGAAATGTAACA 276
QY 318 TTTTGTAACTTAACAACAGATGCAATAGTTTAAATAAATCTTATCTTAACCTTTAAAAA 377
DB 277 CATGATGAATTTGATGCAAGTGCCTTAAATTAATAAATTTGATTTTACATTTTAAAAA 336
QY 378 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404
DB 337 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 363
RESULT 3
ABV57465
ID ABV57465 standard; cDNA; 230 BP.
XX
XX ABV57465;
XX
XX 17-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 57456.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS

```

```
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF
XX XX
XX XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI
XX PI Schlegel R, Endege WO, Monahan JE;
XX XX
XX DR WPI; 2001-662795/76.
XX XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer
XX PS
XX PS Claim 1; Page 11053; 11750pp; English.
XX CC
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ
XX SQ Sequence 230 BP; 101 A; 32 C; 45 G; 52 T; 0 other;
XX
Query Match 14.6%; Score 59.2; DB 23; Length 230;
Best Local Similarity 79.5%; Pred. No. 0.039;
Matches 70; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
XX QY 319 TTGTGTAATCTAACACAGATGCAATAGTTTAAATAAATCTTACTTAACTTTTAAAAAA 378
XX DB 125 TTGAACTTTTCACAGAGATGTAATAATTTTGATAATAAATTTCTTAACCATATAAAAA 184
XX QY 379 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
XX DB 185 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 212
XX
RESULT 4
ABV56953
ID ABV56953 standard; cDNA; 236 BP.
XX AC
XX AC ABV56953;
XX DT
XX DT 17-SEP-2002 (first entry)
XX XX
XX DE Human prostate expression marker cDNA 56944.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX OS
XX PN WO200160860-A2.
XX PD
```

```
XX XX
XX PD 23-AUG-2001.
XX XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI
XX PI Schlegel R, Endege WO, Monahan JE;
XX XX
XX DR WPI; 2001-662795/76.
XX XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer
XX PS
XX PS Claim 1; Page 10969; 11750pp; English.
XX CC
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ
XX SQ Sequence 236 BP; 108 A; 33 C; 40 G; 53 T; 2 other;
XX
Query Match 14.6%; Score 59.2; DB 23; Length 236;
Best Local Similarity 79.5%; Pred. No. 0.039;
Matches 70; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
XX QY 319 TTGTGTAATCTAACACAGATGCAATAGTTTAAATAAATCTTACTTAACTTTTAAAAAA 378
XX DB 126 TTGAACTTTTCACAGAGATGTAATAATTTTGATAATAAATTTCTTAACCATATAAAAA 185
XX QY 379 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
XX DB 186 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 213
XX
RESULT 5
ABV57502
ID ABV57502 standard; cDNA; 574 BP.
XX AC
XX AC ABV57502;
XX DT
XX DT 17-SEP-2002 (first entry)
XX XX
XX DE Human prostate expression marker cDNA 57493.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX OS
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
```

```

XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX DR WPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX PS Claim 1; Page 11059; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX SQ Sequence 574 BP; 241 A; 113 C; 95 G; 124 T; 1 other;

Query Match 14.3%; Score 58; DB 23; Length 574;
Best Local Similarity 58.8%; Pred. NO. 0.064;
Matches 100; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 237 ATTTATTTGCCATAATGAGACCCGTTTTTTATTGTAATCGTCAGTTCCTCAATTAAGTC 296
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 270 ATTTCATTTCACAGGTAATGTTTATATACACTGGCAGCAGCATACATAAACTT 329
OY 297 ATTTTCGACCATACGTAATAATTTTGTATCTAACACAGATGCAATAGTTTAAATAAAC 356
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 330 AGTATGAACTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 389
OY 357 TTATACCTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 390 AAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 439

RESULT 6
ABV58273
ID ABV58273 standard; cDNA; 497 BP.
XX
XX AC ABV58273;
XX
XX DT 13-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 58264.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.

```

```

XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX DR WPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX PS Claim 1; Page 11189; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX SQ Sequence 497 BP; 235 A; 77 C; 69 G; 115 T; 1 other;

Query Match 14.0%; Score 57; DB 23; Length 497;
Best Local Similarity 56.8%; Pred. NO. 0.096;
Matches 105; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 222 ATAGTGAAAATCCGATTATTTCGCATAATGGAGACCCGTTTTTTATTGAATATCGTCAG 281
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 165 ATATATATAAAAAAACCTCTCTGTTATTAGACAAAAAAGCTGTTTTTATTTCGTGAAG 224
OY 282 TTTCCAATTAAGTCATTTTCGAGCCATCTGTAATATTTTGTAACTTAACACAGATGCA 341
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 225 ATATCAATTAATAATAATTTTAGTGGAAATGAAAAAATAAAAAAAAAAAAAA 284
OY 342 ATAGTTTAAATAAACTTATCTTAACCTTTTAAAAAATAAAAAAAAAAAAAA 401
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 285 AAAAAAATAAAAAAAAAAAAAAATAAAAAAAAAAAAAAATAAAAAAAAAAAAAA 344
OY 402 AAAAA 406
DB |||||
DB 345 AAAAA 349

RESULT 7
ABV58662
ID ABV58662 standard; cDNA; 381 BP.
XX
XX AC ABV58662;
XX
XX DT 13-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 58653.
XX

```



KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 11255; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

XX specification of its complement. (I) is useful for:

XX (a) assessing whether a patient is afflicted with prostate cancer;

XX (b) monitoring the progression of prostate cancer in a patient;

XX (c) assessing the efficacy of a test compound to inhibit prostate

XX cancer in a patient;

XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer

XX in a patient;

XX (e) selecting a composition for inhibiting prostate cancer in a patient;

XX (f) assessing the prostate cell carcinogenic potential of a compound;

XX (g) determining whether prostate cancer has metastasized in a patient;

XX (h) assessing the aggressiveness or indolence of prostate cancer in a

XX patient;

XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 381 BP; 196 A; 70 C; 49 G; 66 T; 0 other;

Query Match 13.8%; Score 56; DB 23; Length 381;

Best Local Similarity 77.3%; Pred. No. 0.15;

Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 319 TTGTAACTACACACATGCAATAGTTTAAATAACTTAACTTAACTTTAAAAAAA 378

Db 112 TTGTAACTTTCACAGAGATGTAATAATTTGATAATAAATACTTAACTAATAATCA 171

QY 379 AAAAAAAAAAAAAAAAAAAAAA 406

Db 172 AAAAAAAAAAAAAAAAAAAAAA 199

RESULT 8

ABL32986/C

ID ABL32986 standard; DNA; 17294 BP.

XX ABL32986;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 959.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -

XX Claim 1; SEQ ID NO 959; 32pp + Sequence Listing; German. i

XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention.

XX Sequence 17294 BP; 4922 A; 203 C; 3691 G; 8478 T; 0 other;

Query Match 13.7%; Score 55.8; DB 24; Length 17294;

Best Local Similarity 53.4%; Pred. No. 0.16;

Matches 117; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 184 CTTTTCAGAGGAACTTTCAGACACCTCGCTTTCATAGTGAATCCGATTATT 243

Db 280 CTTTAACTATAAAAAAAAAATAAACCCCTATTAAATCAATTCGATT 221

QY 244 TGCATAATGAGACCGGTTTTTTTATTGAATATCGTCAGTTTCCAATTAAGTCATTTCGA 303

Db 220 TTCTATTAACTCGTACAAATTTTTTTTTTTTTTTTTTTTACTCAATATCAACCA 161

QY 304 GCATATCTGAATATTTTGTAAATCTAACACACATGCAATAGTTTAAATAAATCTTACT 363

Db 160 ACCTAAATAAATTAACTTTCAAAAAATTAATATATAACCATATAAAAAAATAAC 101

QY 364 TAACTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 402

Db 100 TAAATAAAAAAAAAAAAAAAAAAAAAAACAACAA 62

RESULT 9

AAS29049

ID AAS29049 standard; cDNA; 325 BP.

XX AAS29049;

XX 21-NOV-2001 (first entry)

XX cDNA encoding for human DNA-binding protein #20.

XX Human; DNA-binding protein; histone; chromo domain protein;

KW chromatin organisation modifier; y-box binding protein;  
KW DNA organisation; gene transcription; malignant disease;  
KW autoimmune disorder; rheumatic disease; genetic abnormality;  
KW infectious disease; neurological disorder; gene therapy;  
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;  
KW cytostatic; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO20015162-A1.  
PN XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01305.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231143.  
PR 08-SEP-2000; 2000US-0231144.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.

```
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465557/50.
XX P-PSDB; AAU18173.
XX
XX Nucleic acid molecules encoding human secreted chromosomal binding
XX proteins, used in preventing, treating or ameliorating a disorder, e.g.
XX Alzheimer's and Parkinson's diseases and cancers -
XX
XX Claim 4; SEQ ID No 30; 561pp; English.
XX
XX The present invention relates to the isolation of novel DNA-binding
XX proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding
XX for these proteins. DNA-binding proteins such as histones, chromo
XX (chromatin organisation modifier) domain proteins, and Y-box binding
XX proteins may contribute to diseases resulting from aberrant DNA
XX organisation and/or gene transcription. The sequences of the invention
XX are useful in screening assays to identify antagonists and/or agonists
XX that may enhance or block activities mediated by DNA-binding proteins.
XX Blockers of DNA-binding proteins may be useful in treating disorders
XX such as malignant diseases (e.g. cancer), autoimmune disorders
XX (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
XX arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
XX diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
XX disease). The polynucleotide sequences of the invention may also be
XX used in gene therapy. AAS29030-AAS29157 represent cDNA sequences
XX encoding for novel DNA-binding proteins.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 325 BP; 174 A; 38 C; 36 G; 77 T; 0 other;

Query Match 13.68; Score 55.2; DB 22; Length 325;
Best Local Similarity 60.8%; Pred. NO. 0.2;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 259 CCGTTTATTGCAATATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCACTACTGAATAAT 318
    || || || || || || || || || || || || || || || || || || || || || ||
Db 104 CCTTATAATTCACATAAATAAGCAATCTATTAGTCTGATTAGCAATCTAAATGAT 163
    || || || || || || || || || || || || || || || || || || || || || ||

QY 319 TTTGTAATCTAACACAGATGCAATAGTTTAAATAAACCTTATACCTTTTAAAAAAA 378
    || || || || || || || || || || || || || || || || || || || || || ||
Db 164 TCTGTATTAAATGTAATTAAGATTATCTATTGCAAAAAGATATTTTCAACCTTAAAAAAA 223
    || || || || || || || || || || || || || || || || || || || || || ||

QY 379 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 406
    || || || || || || || || || || || || || || || || || || || || || ||
Db 224 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 251
    || || || || || || || || || || || || || || || || || || || || || ||

RESULT 10
AAS29124
ID AAS29124 standard; cDNA; 621 BP.
XX
XX AC AAS29124;
XX
XX 21-NOV-2001 (first entry)
XX
XX cDNA encoding for human DNA-binding protein #95.
XX
XX Human; DNA-binding protein; histone; chromo domain protein;
XX chromatin organisation modifier; Y-box binding protein;
XX DNA organisation; gene transcription; malignant disease;
XX autoimmune disorder; rheumatic disease; genetic abnormality;
XX infectious disease; neurological disorder; gene therapy;
```

```
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
XX cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO20015162-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01305.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0228668.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 21-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
```



```
XX PN EPI146052-A2.
XX PD 17-OCT-2001.
XX PF 12-APR-2001; 2001EP-0303488.
XX PR 14-APR-2000; 2000US-197279P.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX XX Presnail JK, Weng Z, Wong JF;
XX PI WPI: 2002-001056/01.
XX DR P-PSDB; AAG77934.
XX XX
XX PT New polynucleotides encoding Arthropod defensin polypeptides for
XX PT improving the microbial resistance of plants and animals by gene
XX PT therapy -
XX XX
XX PS Claim 6; Page 19; 22pp; English.
XX CC The sequence represents a novel Arthropod defensin from Arioipe sp.
XX CC clone atic.pk0. The invention relates to novel polynucleotides encoding
XX CC Arthropod defensin polypeptides. The polypeptides and polynucleotides of
XX CC the invention have antimicrobial activity. The polynucleotides may be
XX CC used to genetically engineer cells and organisms, especially plants, to
XX CC alter their resistance to a wide range of microbial pathogens. The
XX CC polynucleotides may also have a use in gene therapy.
XX SQ Sequence 351 BP; 95 A; 84 C; 69 G; 103 T; 0 other;
Query Match 13.4%; Score 54.6; DB 24; Length 351;
Best Local Similarity 54.9%; Pred. No. 0.26;
Matches 130; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
QY 23 TTCTCCACTCAGCTTCAGAAATGAATCCATAGTATATTTTCATCGTTCCTTTGTTGCC 82
Db ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7 TCTGTGCGACATTTCCAAAAAATGAATCGGAGAGTCTGTGTGTGATCTGCTAGT---C 63
QY 83 TTCTGTATTTGGAGGATGGATGTAGAACTGTTTGGATGTCCTCTTTAATGCAGGA 142
Db ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 GTCGTGCTTTTGGCACAGTGCACCGGTGAAGTGTGTTCGGCTGCCCTTCGACCATG 123
QY 143 AAATGCCATAGACATTCGAAAAGTATTCGTCGTAGAGAGGCTTTTGCAGAGGAACATTC 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 CAGTGTCAATCATTCGAGGAGCATCAATATAGGGGAGGATAGTGCACCAACTTATTC 183
QY 203 AGGACAACCTGGTTCGTATAGGTGAAATCCGATTTATTTGCCATATGAGACC 259
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 AAGCGACCTGCAAGTGTACGGATGATGACCCCTCTCTCAGACAGGAGGCC 240
RESULT 12
AAS60450/c
ID AAS60450 standard; cDNA: 425 BP.
XX AC AAS60450;
XX XX
XX DT 29-JAN-2002 (first entry)
XX DE Human cancer agent-sensitive marker #181.
XX DE
XX KW Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
XX KW squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
XX KW lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
XX KW Hodgkin's disease; glioma; ss.
XX OS Homo sapiens.
XX XX
XX PN WO200179556-A2.
XX XX
XX PD 25-OCT-2001.
```

```
XX 13-APR-2001; 2001WO-US12132.
XX PF
XX PR 14-APR-2000; 2000US-197538P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX XX Lillie J, Brown JL, Bolt A, Van Huffel C;
XX DR WPI: 2001-602933/68.
XX XX
XX PT Novel nucleic acid, used as a marker to determine the effectiveness of
XX PT using TAXOL to treat cancer cell growth in individuals -
XX PS Claim 1; Page 221; 527pp; English.
XX XX
XX CC The invention relates to 1046 novel nucleic acids which are used as
XX CC markers for determining the sensitivity of a cancer cell to the
XX CC anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
XX CC they are shown to express one of the 242 sensitivity markers or the
XX CC cells are shown not to express one of the 804 resistance markers.
XX CC The methods can be used to determine the effectiveness of TAXOL
XX CC in the treatment of cancer cell growth in an individual. The markers
XX CC can be used as targets in developing anti-cancer agents such as
XX CC chemotherapeutic compounds. The markers can also be used as targets in
XX CC developing treatments for cancer, particularly those cancers which
XX CC display resistance to agents and exhibit expression of the markers. The
XX CC anticancer agents developed by the novel method can be used to treat
XX CC cancer. Probes based on the markers can be used to detect transcripts or
XX CC genomic sequences corresponding to the markers, in the identification of
XX CC cells or tissues which mis-express the protein. Cancers which may
XX CC be targeted include carcinoma (e.g. squamous cell carcinoma),
XX CC sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
XX CC lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
XX CC tumours (e.g. glioma). The present sequence is one of the 1046
XX CC novel cancer cell markers.
XX SQ Sequence 425 BP; 159 A; 41 C; 28 G; 194 T; 3 other;
Query Match 13.4%; Score 54.6; DB 22; Length 425;
Best Local Similarity 53.4%; Pred. No. 0.26;
Matches 111; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 199 TTTCAGACACACCTGCGTTTGTCTATAGGTGAAATCGGATTTATTTGCCATATGAGAC 258
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 TTTTAAAAAAGACGGGGGGGAAAAAAGAAACCTTTTTTTTTTTTGGGGGAA 156
QY 259 CCGTTTTTATTGAATATCGTCAGTTTCCCAATTAAGTCATTCGAGCCATCTGAATAAT 318
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 ATTTTTTTTTTTTTTTTTTTTTTTTAAAAATTTTTTTTTTTTNNNAAAAAAGAAAAA 96
QY 319 TTGTGAATCTAACACAGATGCAATAGTTTAAATAAACTTATATTAACCTTTTAAAAA 378
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 36
QY 379 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 8
RESULT 13
ABQ54618
ID ABQ54618 standard; cDNA: 2736 BP.
XX AC ABQ54618;
XX XX
XX DT 22-AUG-2002 (first entry)
XX DE Human ovarian antigen HDPN96 cDNA, SEQ ID NO:498.
XX DE
XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
XX KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
XX KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
```

KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; chromosome 1q22;  
 KW gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200200677-A1.  
 PN  
 XX  
 XX 03-JAN-2002.  
 PD  
 XX  
 XX 07-JUN-2001; 2001WO-US18569.  
 PF  
 XX  
 XX 07-JUN-2000; 2000US-209467P.  
 PR  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 XX WPI; 2002-147878/19.  
 DR  
 XX P-PSDB; ABP41541.  
 DR  
 XX  
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 XX Claim 1; SEQ ID No 498; 2922pp; English.

PS The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompassing polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2736 BP; 989 A; 528 C; 543 G; 672 T; 4 other;  
 Query Match 13.48; Score 54.6; DB 24; Length 2736;  
 Best Local Similarity 48.78; Pred. No. 0.26;  
 Matches 147; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

Qy 105 TTGTAGAGCTGGTTTGGATGTCCTTTAATGCAGGAAATGCATAGACATGCCAAA 164  
 |||||  
 Db 2432 TTCTAAATCTGGATGTCATTGACGATTAATGTTATGGAGATAGGCTAAGTCCTTAA 2491

Qy 165 GTATTGCTGTAGAGAGGCTTTTGCAGAGAACTTTTCAGGACAACTCGCTTGCTATA 224  
 |||||  
 Db 2492 AAATGTACATATACCTGGTTGAATATACAACTATACATACACACCATATATACTA 2551  
 |||||  
 Qy 225 GTGAAATCCGATTTTATTTGGCCATAATGGAGACCGCTTTTATTGAATATCGTCAGTTT 284  
 |||||  
 Db 2552 GCTGTTAATCCTATGGAATGGGTATTGGGAGTGCCTTTTAAATTTTTCATAGTTTTT 2611  
 |||||  
 Qy 285 CCAATTAAAGTCATTTTCGAGCCATACCTACTGAATTAATTTTGAATCTAACACAGATGCAATA 344  
 |||||  
 Db 2612 TTAATAAAATGGCATATTTTGCATCTTACAACTTCTATAATTTGAAAAAATAAATAACAT 2671  
 |||||  
 Qy 345 GTTAAATAAACTTATACCTTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404  
 |||||  
 Db 2672 TATCTTTTGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2731  
 |||||  
 Qy 405 AA 406  
 ||  
 Db 2732 AA 2733

RESULT 14  
 ABV37528/c  
 ID ABV37528 standard; cDNA; 372 BP.

XX AC ABV37528;  
 XX DT 16-SEP-2002 (first entry)  
 XX DE Human prostate expression marker cDNA 37519.  
 XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX OS Homo sapiens.  
 XX PN WO200160860-A2.  
 XX PD 23-AUG-2001.  
 XX PF 20-FEB-2001; 2001WO-US05171.  
 XX PR 17-FEB-2000; 2000US-183319P.  
 XX PR 16-MAR-2000; 2000US-189862P.  
 XX PR 25-MAY-2000; 2000US-207454P.  
 XX PR 09-JUN-2000; 2000US-211314P.  
 XX PR 18-JUL-2000; 2000US-219007P.  
 XX PR 13-DEC-2000; 2000US-255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX PI Schlegel R, Endege WO, Monahan JE;  
 XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 7697-7698; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 372 BP; 141 A; 37 C; 50 G; 144 T; 0 other;

Query Match 13.4%; Score 54.4; DB 23; Length 372;  
Best Local Similarity 57.7%; Pred. No. 0.28; Indels 0; Gaps 0;  
Matches 97; Conservative 0; Mismatches 71;  
QY 239 TTATTGGCCATAATGGAGACCGTTTATTGAATATATCGTCAGTTTCCAAATTAAGTCAT 298  
DB 295 TTTTATTTTATTTTAAATACATTTTAAATTCCTTTTAAAGCCCATTTATTTTTC 236  
QY 299 TTCGAGCCACTACTGATAATTTTGAATCTAACACAGATGCAATAGTTTAATAAAGCTT 358  
DB 235 TTTTAATAATTTTATTTTATTTTAAATATACCCCTTTTATTTTAAATAAATTT 176  
QY 359 ATACTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 406  
DB 175 TTTTTCGTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 128

RESULT 15  
ABV10351/c  
ID ABV10351 standard; cDNA; 447 BP.  
XX  
AC ABV10351;  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 10342.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 1663; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 447 BP; 98 A; 75 C; 91 G; 183 T; 0 other;

Query Match 13.4%; Score 54.4; DB 23; Length 447;  
Best Local Similarity 65.8%; Pred. No. 0.28; Indels 0; Gaps 0;  
Matches 79; Conservative 0; Mismatches 41;  
QY 287 AATTAAAGTCATTTTCGAGCCATCTGATAATTTTGAATCTTAACAACAGATGCAATAGT 346  
DB 428 AATAAATAAATTTAAATAAACTTAAATATATAATCAATTAATATAATCAAAAA 369  
QY 347 TTAATAAAGTTTAACTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406  
DB 368 TAAATAAAGTTTAACTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 309

Search completed: May 5, 2003, 19:18:44  
Job time : 238 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 19:18:48 ; Search time 50 seconds  
(without alignments)  
2490.215 Million cell updates/sec

Title: US-09-829-481-3  
Perfect score: 406  
Sequence: 1 cctctactaactcactaagt.....aaaaaaaaaaaaaaaaaaaaa 406

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A-COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B-COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A-COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B-COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS-COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55.6	13.7	1652	4	US-09-627-650B-2
2	55.6	13.7	1652	4	US-09-436-063C-2
3	52.8	13.0	3581	2	US-08-738-349-1
4	52.4	12.9	958	2	US-08-757-046A-5
5	52.4	12.9	958	3	US-09-447-208-5
6	52.4	12.9	958	3	US-09-135-988-5
7	52.4	12.9	958	4	US-09-277-716-5
8	52.4	12.9	958	4	US-08-597-274A-5
9	52.4	12.9	958	4	US-08-908-909-5
10	52.4	12.9	958	4	US-09-609-161B-5
11	52.4	12.9	958	4	US-08-990-103-5
12	52	12.8	857	1	US-08-308-883-1
13	52	12.8	857	1	US-08-730-163-1
14	52	12.8	857	4	US-08-256-799-1
15	52	12.8	857	4	US-08-462-437-1
16	50.8	12.5	1582	3	US-08-545-196B-10
17	50.8	12.5	1582	3	US-08-545-196B-12
18	50.6	12.5	1411	4	US-08-964-127-5
19	50.6	12.5	1411	4	US-09-496-692-5
20	50.6	12.5	1641	1	US-08-300-903A-8
21	50.4	12.4	2360	3	US-08-836-567-9
22	50.4	12.4	3927	4	US-09-293-238B-1
23	50.4	12.4	3933	1	US-08-199-776-1
24	50.4	12.4	3933	3	US-08-663-731-1
25	50.4	12.4	3933	3	US-08-879-338-1
26	50.4	12.4	3933	5	PCT-US95-02044-1
27	50.2	12.4	1474	4	US-08-821-994-64

28	50.2	12.4	2790	3	US-08-800-291B-1
29	50	12.3	1522	4	US-09-413-574-1
30	49.6	12.2	2671	6	5168051-9
31	49.4	12.2	458	1	US-08-524-757-1
32	48.4	11.9	841	4	US-09-004-731-40
c 33	48.4	11.9	841	4	US-09-004-731-42
34	48.4	11.9	841	4	US-09-032-215-46
c 35	48.4	11.9	841	4	US-09-032-215-48
36	48.4	11.9	841	4	US-08-749-699-40
c 37	48.4	11.9	841	4	US-08-749-699-42
38	48.4	11.9	841	4	US-09-004-729-40
c 39	48.4	11.9	841	4	US-09-004-729-42
40	48.4	11.9	1558	1	US-08-455-550-7
41	48.4	11.9	1736	3	US-09-182-816-22
c 42	48.4	11.9	1736	3	US-09-182-816-24
43	48.4	11.9	1736	3	US-09-471-528-22
c 44	48.4	11.9	1736	3	US-09-471-528-24
45	48.4	11.9	1736	4	US-09-634-530-22

ALIGNMENTS

RESULT 1  
US-09-627-650B-2  
; Sequence 2, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; TITLE OF INVENTION: Neuromuscular Junction GABA Receptors and  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1652  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-2

Query Match	13.7%	Score 55.6;	DB 4;	Length 1652;
Best Local Similarity	55.1%	Pred. NO. 0.0017;		
Matches 109;	Conservative 0;	Mismatches 89;	Indels 0;	Gaps 0;
QY 209	ACCTCGCTTTCGTATAGGTGAAATCCGATTTTATTTGCCATAATCGACCGGTTTAT 268			
Db 1450	ACCTTCATTAATCTCAATCCAACTTCCATCATCTTCCATTCGATATCTCTTTCT 1509			
QY 269	TGAATATCGTCAGTTTCCAAATTAAGTCATTTTCGACCATCTGAATATTTGTAATCT 328			
Db 1510	TCCACAGAAGCCTTTTTCGTTTTTTTATTTGATTTATTTACGGATTTTAGATAAT 1569			
QY 329	AACAACAGATGCAATAGTTTAAATAAATACTTACTTAACTTTAAAAAATAAAAAA 388			
Db 1570	GCACAGATGCCCTTCATTCGTCATAATAATTTATTTTAAATGTCGAAAAAATAAAAA 1629			
QY 389	AAAAAAAAAAAAAAAAAAAA 406			
Db 1630	AAAAAAAAAAAAAAAAAAAA 1647			
RESULT 2				
US-09-436-063C-2				
; Sequence 2, Application US/09436063C				
; Patent No. 6407210				
; GENERAL INFORMATION:				

```
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-2

Query Match          13.7%; Score 55.6; DB 4; Length 1652;
Best Local Similarity 55.1%; Pred. No. 0.0017;
Matches 109; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 209 ACCTGCGTTTGTCTATAGTGAAATCCGATTATTTGCCATAATGGAGACCGGTTTAT 268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1450 ACCTTCATTATCAATCAACATCTCTCATCATTTTCCATTCGAATATCTCTTTTCT 1509
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 269 TGAATATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCATACGAATAATTTGTATCT 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1510 TGCACAGAGCGCTTTTTCGTTTTTTTATTTGATTTATTTTACGGGATTTTATAGATAAT 1569
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 329 AACACAGATGCAATAGTTTAAATAAAGCTTATCTTAACTTTTAAATAAAAAAAAAA 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1570 GCACAGATGCTCATCTGCTCAATAAATTTATTTAATGTGCAAAAAAAAAAAAAA 1629
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 389 AAAAAAAAAAAAAAAAAA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1630 AAAAAAAAAAAAAAAAAA 1647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-738-349-1
; Sequence 1, Application US/08738349
; Patent No. 5869638
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Kawai, Shinji
; APPLICANT: Tsujimura, Atsushi
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and
; FILE OF INVENTION: Process for Its Production
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,349
; FILING DATE: 25-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/364,439
; FILING DATE:
; APPLICATION NUMBER: US 08/112,061
; FILING DATE: 26-AUG-1993
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Barker, M. P.
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 02481.1323-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; STRAIN: osteoblastic cell line MC3T3E1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 284..2671
US-08-738-349-1

Query Match          13.0%; Score 52.8; DB 2; Length 3581;
Best Local Similarity 71.9%; Pred. No. 0.0071;
Matches 69; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 311 TGAATAATTTTGTATCTTAACAACAGATGCAATAGTTTAAATAAACTTATCTTAAC 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3443 TGCTTTAATATGAGTTCATATATAAGAACGACCTTGAATAATAAAAAAGATCTTTT 3502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 371 TAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3503 TAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3538
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-757-046A-5
; Sequence 5, Application US/08757046A
; Patent No. 5876995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,046A
; FILING DATE: 11-25-96
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,274
; FILING DATE: 02-06-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6680-105B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
```





```
;
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaquorin-encoding gene
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: 5,093,240
; AUTHORS: Inouye et al.
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 82
; PAGES: 3154-3158
; DATE: (1985)
;
; US-08-597-274A-5

Query Match 12.9%; Score 52.4; DB 4; Length 958;
Best Local Similarity 57.2%; Pred. No. 0.0072;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 241 ATTTGCCATAATGGAGACCCGTTTTTATGTAATATCGTCAGTTTCCCAATTAAGTCATTT 300
Db 779 ATTTCCAAATTTTGAACGATTTCAATCGTTGTGTTGATTTTGTCTAATAGGAACAGA 838

QY 301 CGAGCCATACTCAATAATTTTCTAATCTAACACAGATGCAATAGTTTAAATAAATCTAT 360
Db 839 TTAATCGAATGATTAGTTGTTTATTAATCAACAGAACTTACAATCGAAATGTAATAA 898

QY 361 ACTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
Db 899 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944

RESULT 9
US-08-908-909-5
; Sequence 5, Application US/08908909
; Patent No. 6416960
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: DETECTION AND VISUALIZATION OF
; TITLE OF INVENTION: NEOPLASTIC TISSUES AND OTHER TISSUES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/908,909
; FILING DATE: 08-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/023,374
; FILING DATE: 08-AUG-1996
; CLASSIFICATION: 435
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6680-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaquorin-encoding gene
; PUBLICATION INFORMATION:
; AUTHORS: Inouye et al.
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 82
; PAGES: 3154-3158
; DATE: (1985)
;
; DOCUMENT NUMBER: 5,093,240
;
; US-08-908-909-5

Query Match 12.9%; Score 52.4; DB 4; Length 958;
Best Local Similarity 57.2%; Pred. No. 0.0072;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 241 ATTTGCCATAATGGAGACCCGTTTTTATGTAATATCGTCAGTTTCCCAATTAAGTCATTT 300
Db 779 ATTTCCAAATTTTGAACGATTTCAATCGTTGTGTTGATTTTGTCTAATAGGAACAGA 838

QY 301 CGAGCCATACTCAATAATTTTCTAATCTAACACAGATGCAATAGTTTAAATAAATCTAT 360
Db 839 TTAATCGAATGATTAGTTGTTTATTAATCAACAGAACTTACAATCGAAATGTAATAA 898

QY 361 ACTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
Db 899 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944

RESULT 10
US-09-609-161B-5
; Sequence 5, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROMUNE, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS,
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 5
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Aequorea (luminescent jellyfish)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)..(702)
; OTHER INFORMATION: Apoaeguorin-encoding gene
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: US 07/105,602
; PATENT FILING DATE: 1987-10-08
; PUBLICATION DATE: 1992-03-03
; PUBLICATION INFORMATION:
; AUTHORS: Inouye, S.
; AUTHORS: No. 6436682uchi, M.
; AUTHORS: Sakaki, Y.
; AUTHORS: Takagi, Y.
; AUTHORS: Miyata, T.
; AUTHORS: Iwanaga, S.
; AUTHORS: Miyata, T.
; AUTHORS: Tsuji, F.I.
; TITLE: Cloning and sequence analysis of cDNA for the luminescent protein
; TITLE: aequorin
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 82(10)
; PAGES: 3154-3158
; DATE: 1985-05
; US-09-609-161B-5

Query Match 12.9%; Score 52.4; DB 4; Length 958;
Best Local Similarity 57.2%; Pred. No. 0.0072;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 241 ATTTGCCATAATGAGACCGCGTTTATTGAATATCGTCAGTTTCCAATTAAGTCATT 300
Db 779 ATTTTCCAAATTTTGAACGATTCATCGTTTGTGTTGATTTTGTAAATAGGAACAGA 838

QY 301 CGAGCCATCTGAATAATTTTGTATCTAACACAGATGCAATGTTTAAATAACTTAT 360
Db 839 TTAATCGAATGATGTGTTTATTCATCAACAGAACTTACAATCGAAAAAGTAAAA 898

QY 361 ACTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
Db 899 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944

RESULT 11
US-09-990-103-5
; Sequence 5, Application US/08990103
; Patent No. 6458547
; GENERAL INFORMATION:
; APPLICANT: Bruce J. Bryan
; APPLICANT: Stephen Gaalema
; APPLICANT: Randall B. Murphy
; TITLE OF INVENTION: APPARATUS AND METHOD FOR DETECTING AND
; TITLE OF INVENTION: IDENTIFYING INFECTIOUS AGENTS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,103
; FILING DATE: 12-DEC-1997
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,675,
; FILING DATE: 02-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,745
; FILING DATE: 12-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6680-112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaeguorin-encoding gene
; PUBLICATION INFORMATION:
; AUTHORS: Inouye et al.
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 82
; PAGES: 3154-3158
; DATE: (1985)
; DOCUMENT NUMBER: 5,093,240
; US-08-990-103-5

Query Match 12.9%; Score 52.4; DB 4; Length 958;
Best Local Similarity 57.2%; Pred. No. 0.0072;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 241 ATTTGCCATAATGAGACCGCGTTTATTGAATATCGTCAGTTTCCAATTAAGTCATT 300
Db 779 ATTTTCCAAATTTTGAACGATTCATCGTTTGTGTTGATTTTGTAAATAGGAACAGA 838

QY 301 CGAGCCATCTGAATAATTTTGTATCTAACACAGATGCAATGTTTAAATAACTTAT 360
Db 839 TTAATCGAATGATGTGTTTATTCATCAACAGAACTTACAATCGAAAAAGTAAAA 898

QY 361 ACTTAACCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
Db 899 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944

RESULT 12
US-08-308-883-1
; Sequence 1, Application US/08308883
; Patent No. 5576300
; GENERAL INFORMATION:
; APPLICANT: Mukerji, P.
; APPLICANT: Prieto, P. A.
; APPLICANT: Seo, A. E.-Y.
; APPLICANT: Baxter, J. H.
; APPLICANT: Cummings, R.D.
; TITLE OF INVENTION: Method for Inhibition of Human Rotavirus Infection.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lonnie R. Drayer
; ADDRESSEE: Ross Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
```

```
STATE: Ohio
COUNTRY: United States
ZIP: 43215
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh System 7.1
SOFTWARE: ClarisWorks 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308.883
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA: NO. 5576300 applicable
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-3774
TELEFAX: (614) 624-3074
INFORMATION FOR SEQ ID NO: 1:
TELEX: No. 5576300e
SEQUENCE CHARACTERISTICS:
LENGTH: 857 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: cDNA
DESCRIPTION: Human milk kappa-casein
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: Human
ORGANISM: Homo sapiens
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: Adult
HAPLOTYPE:
TISSUE TYPE: Mammary gland
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Mammary Gland
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY: CDS
LOCATION: 45...593
IDENTIFICATION METHOD: DNA sequencing and restriction analysis
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human
PUBLICATION INFORMATION:
AUTHORS: L. Hansson et al
TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER: PCT/WO93/15196
FILING DATE: 25-JAN-1993
PUBLICATION DATE: 05-AUG-1993
RELEVANT RESIDUES IN SEQ ID NO:
US-08-308-883-1
Query Match 12.8%; Score 52; DB 1; Length 857;
Best Local Similarity 67.6%; Pred. No. 0.0086;
Matches 73; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 299 TTCGAGCCATCTGAATATTTTGTAACTCAACACAGATGCAATGTTTAAATAAACTT 358
DB 741 TTCATGCCACATTCATATTTTGTCTTGCACATAAAGCCCACTGATTCGCAAAAAA 800
QY 359 ATACTTAACTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
```

Db 801 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 848

RESULT 13  
US-08-730-163-1  
Sequence 1, Application US/08730163  
Patent No. 5712250  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R.D.  
TITLE OF INVENTION: Product for Inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: Ross Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/730.163  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/308.882  
FILING DATE: 16-SEP-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: No. 5712250e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 857 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human milk kappa-casein  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: Human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult  
HAPLOTYPE:  
TISSUE TYPE: Mammary gland  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: Human Mammary Gland  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45...593  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER: PCT/WO93/15196  
FILING DATE: 25-JAN-1993  
PUBLICATION DATE: 05-AUG-1993  
RELEVANT RESIDUES IN SEQ ID NO:

;; AUTHORS: L. Hansson et al  
;; TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The  
;; JOURNAL:  
;; VOLUME:  
;; ISSUE:  
;; PAGES:  
;; DATE:  
;; DOCUMENT NUMBER: PCT/NO93/15196  
;; FILING DATE: 25-JAN-1993  
;; PUBLICATION DATE: 05-AUG-1993  
;; RELEVANT RESIDUES IN SEQ ID NO:  
US-08-730-163-1

Query Match 12.8%; Score 52; DB 1; Length 857;  
Best Local Similarity 67.6%; Pred. No. 0.0086;  
Matches 73; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 299 TTCGAGCCACTGTAATAATTTTGTAACTCAACAACAGATGCAATAGTTTAAATAAACTT 358  
Db 741 TTCATGCCACATTCATATTTTGTATTCGACATAAAGCCCACTGATTCGCAAAAAA 800  
QY 359 ATACITTAACITTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 406  
Db 801 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 848

RESULT 14  
US-08-256-799-1  
; Sequence 1, Application US/08256799  
; Patent No. 6222094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olle  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,799  
; FILING DATE: 06-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO

;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 45...593  
;; FEATURE:  
;; NAME/KEY: mat\_peptide  
;; LOCATION: 45...593  
;; FEATURE:  
;; NAME/KEY: sig\_peptide  
;; LOCATION: 45...104  
;; FEATURE:  
;; NAME/KEY: 5'UTR  
;; LOCATION: 13...44  
;; FEATURE:  
;; NAME/KEY: 3'UTR  
;; LOCATION: 594...848  
US-08-256-799-1

Query Match 12.8%; Score 52; DB 4; Length 857;  
Best Local Similarity 67.6%; Pred. No. 0.0086;  
Matches 73; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 299 TTCGAGCCACTGTAATAATTTTGTAACTCAACAACAGATGCAATAGTTTAAATAAACTT 358  
Db 741 TTCATGCCACATTCATATTTTGTATTCGACATAAAGCCCACTGATTCGCAAAAAA 800  
QY 359 ATACITTAACITTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 406  
Db 801 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 848

RESULT 15  
US-08-462-437-1  
; Sequence 1, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olle  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:



Tue May 6 09:32:27 2003

LENGTH: 857 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45..593  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 45..593  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 45..104  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 13..44  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 594..848  
US-08-462-437-1

Query Match 12.8%; Score 52; DB 4; Length 857;  
Best Local Similarity 67.6%; Pred. No. 0.0086;  
Matches 73; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
QY 299 TTCGAGCCATCTGAATAATTTTGTAACTCAACAACAGATGCAATAGTTTAAATAAATT 358  
Db 741 TTCATGCCACATTCAATATTTTGATTCTTGCACAATAAAGCCCAACTGATTGCAAAAAA 800  
QY 359 ATACTTAACTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 406  
Db 801 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 848

Search completed: May 5, 2003, 20:26:55  
Job time : 54 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 19:53:13 : Search time 143 Seconds  
(without alignments)  
3354.812 Million cell updates/sec

Title: US-09-829-481-3  
Perfect score: 406  
Sequence: 1 ctctactacatactactaagt.....aaaaaaaaaaaaaaaaaaaaa 406

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues  
Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	100.0	406	10	US-09-829-481-3
2	76	18.7	386	10	US-09-829-481-5
3	55.2	13.6	325	9	US-10-091-483-30
4	55.2	13.6	325	9	US-09-764-846-30
5	55.2	13.6	621	9	US-10-091-483-105
6	55.2	13.6	621	10	US-09-764-846-105
7	54.6	13.4	351	10	US-09-829-481-7
c 8	54.6	13.4	425	10	US-09-834-975-451
c 9	53.8	13.3	380	10	US-09-960-352-9335
10	53.8	13.3	461	10	US-09-829-481-1
11	53.6	13.2	471	9	US-09-918-995-14052
c 12	53.2	13.1	236	10	US-09-960-352-12183
c 13	53	13.1	397	10	US-09-960-352-13784
14	52.8	13.0	393	10	US-09-960-352-5187
15	52.8	13.0	454	9	US-09-918-995-13971
16	52.6	13.0	3293	10	US-09-764-864-123
17	52.4	12.9	958	9	US-10-126-139-5
18	52.4	12.9	958	9	US-10-126-798-5
19	52.4	12.9	958	10	US-09-803-211-5

c 20	52.4	12.9	958	10	US-09-746-485A-5	Sequence 5, Appli
21	52.2	12.9	368	10	US-09-834-975-47	Sequence 47, Appli
22	52.2	12.9	426	9	US-09-918-995-17732	Sequence 17732, A
23	52.2	12.9	432	9	US-09-918-995-13651	Sequence 13651, A
24	52.2	12.9	1387	9	US-10-091-483-113	Sequence 113, App
25	52.2	12.9	1387	10	US-09-764-846-113	Sequence 113, App
26	52.2	12.9	3957	10	US-09-764-853-77	Sequence 77, Appl
27	52	12.8	857	9	US-09-954-531-977	Sequence 70, Appl
28	52	12.8	1499	9	US-10-114-893-70	Sequence 15258, A
29	51.8	12.8	219	10	US-09-878-574-15258	Sequence 93, Appl
30	51.8	12.8	1457	9	US-09-989-920-93	Sequence 8, Appli
31	51.8	12.8	4000	9	US-09-981-353-8	Sequence 6976, Ap
c 32	51.6	12.7	337	10	US-09-960-352-6976	Sequence 3371, Ap
c 33	51.6	12.7	347	10	US-09-960-352-3371	Sequence 196, App
34	51.4	12.7	1772	9	US-10-102-806-196	Sequence 5559, Ap
c 35	51	12.6	358	10	US-09-960-352-5559	Sequence 15014, A
c 36	50.8	12.5	375	10	US-09-960-352-15014	Sequence 53, Appl
c 37	50.8	12.5	396	9	US-09-970-966-53	Sequence 53, Appl
c 38	50.8	12.5	396	10	US-09-825-294-53	Sequence 53, Appl
39	50.8	12.5	1001	9	US-09-986-480-93	Sequence 5, Appli
40	50.6	12.5	1411	9	US-10-000-273-5	Sequence 39, Appl
41	50.6	12.5	3239	9	US-10-174-363-39	Sequence 13, Appl
42	50.4	12.4	2511	9	US-10-091-752A-13	Sequence 668, App
43	50.2	12.4	285	10	US-09-834-975-668	Sequence 1243, Ap
c 44	50.2	12.4	291	10	US-09-960-352-1243	Sequence 7994, Ap
45	50.2	12.4	373	10	US-09-960-352-7994	

ALIGNMENTS

RESULT 1  
US-09-829-481-3  
; Sequence 3, Application US/09829481  
; Patent No. US20020069427A1  
; GENERAL INFORMATION:  
; APPLICANT: Presnail, James  
; APPLICANT: Wong, Zude  
; TITLE OF INVENTION: Arthropod Defensins  
; FILE REFERENCE: BB1441 US NA  
; CURRENT APPLICATION NUMBER: US/09/829,481  
; CURRENT FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: 60/197279  
; PRIOR FILING DATE: 2000-04-14  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 3  
; LENGTH: 406  
; TYPE: DNA  
; ORGANISM: Vaejovis carolinianus  
US-09-829-481-3

Query Match	100.0%	Score 406:	DB 10:	Length 406:
Best Local Similarity	100.0%	Pred No. 1.8e-74;		
Matches 406:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	CTCTACTACAATCACTAAGTTCTTTCTCCACTCAGCTTCAAGAATGAATCCATAGCTAT	60		
Db 1	CTCTACTACAATCACTAAGTTCTTTCTCCACTCAGCTTCAAGAATGAATCCATAGCTAT	60		
Qy 61	TATTTTCATCCTTCTTGTTCCTCTCTATTTTGGAGATGGATTTGTAGAACTGGTTT	120		
Db 61	TATTTTCATCCTTCTTGTTCCTCTCTATTTTGGAGATGGATTTGTAGAACTGGTTT	120		
Qy 121	TGGATGTCCTTCTTATCGAGCAAAATGCCATAGACATTTGCAAAAGTATTCCTCGTAGAGG	180		
Db 121	TGGATGTCCTTCTTATCGAGCAAAATGCCATAGACATTTGCAAAAGTATTCCTCGTAGAGG	180		
Qy 181	AGGCTTTTGCAGAGGAACCTTTTCAGGACAACCTGCTTTCCTATAGGTGAAATCCGATTT	240		
Db 181	AGGCTTTTGCAGAGGAACCTTTTCAGGACAACCTGCTTTCCTATAGGTGAAATCCGATTT	240		

```
QY 241 ATTGCGCATAATGGAGACCGGTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTT 300
Db 241 ATTGCGCATAATGGAGACCGGTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTT 300
QY 301 CGAGGCATACCTGAATAATTTTCTAATCTAACACAGATGCAATAGTTTAAATAAACTTAT 360
Db 301 CGAGGCATACCTGAATAATTTTCTAATCTAACACAGATGCAATAGTTTAAATAAACTTAT 360
QY 361 ACTTAACTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406
Db 361 ACTTAACTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406

RESULT 2
US-09-829-481-5
; Sequence 5, Application US/09829481
; Patent No. US20020069427A1
; GENERAL INFORMATION:
; APPLICANT: Presnail, James
; APPLICANT: Wong, Zude
; APPLICANT: Wong, James
; TITLE OF INVENTION: Arthropod Defensins
; FILE REFERENCE: BB1441 US NA
; CURRENT APPLICATION NUMBER: US/09/829,481
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/197279
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 386
; TYPE: DNA
; ORGANISM: Argiope sp.
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (351)
; NAME/KEY: unsure
; LOCATION: (364)..(365)
; NAME/KEY: unsure
; LOCATION: (386)
US-09-829-481-5

Query Match 18.7%; Score 76; DB 10; Length 386;
Best Local Similarity 54.4%; Pred. No. 9e-07;
Matches 178; Conservative 0; Mismatches 141; Indels 8; Gaps 1;

QY 78 TTGCTTCGTGATTTTGGAGATGGATGTTAGAGCTGGTTTTGGATGCCCTTTTAATG 137
Db 45 TAGTCGCTGTGCTTTTGGCCACAGTGGCCGCGGGAAGCTGTTTCGGCTGCCCTTCGACC 104
QY 138 CAGGAAATGCCATAGACATTCGAAAAGTATTCGTCGTAGAGAGGCGTTTTCGACAGGAA 197
Db 105 AGATGCAGTGTCAATCATTTGCAGGAGCATCAATACAGGGGAGGATCTGCACCACT 164
QY 198 CTTTCAGACAACTCGGTTTGCATAGGTGAAATCCGATTTATTTGCCATAATGGAGA 257
Db 165 TATTCAAGCGCACCTGCAAGTGTTCACGATGATGACCCCTCCCTCTCAGACAGGAGC 224
QY 258 CCGCTTTTATTGAATTCGTCAGTTTCCCAATTAAGTCATTTTCGAGCCATCTGAATAA 317
Db 225 CCACACCTT-----TCATTGACATCCGATTCGGATTTTCCAAATGCAAAATGTAA 276
QY 318 TTTTGTAACTTAACAACAGATGCAATAGTTTAAATAAATCTATATCTAACTTTTAAAAA 377
Db 277 CATGATGAATTTGATGCAAGTCCTTAATTTAAATAAATTTGATTTTACATTTTAAAAA 336
QY 378 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 404
Db 337 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 363
```

RESULT 3  
US-10-091-483-30

```
; Sequence 30, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-483-30
```

```
Query Match 13.6%; Score 55.2; DB 9; Length 325;
Best Local Similarity 60.8%; Pred. No. 0.016;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 259 CCGTTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCATCTGAATAT 318
Db 104 CCTTTATAATTCACATAAATAAAGCATCTATTAGTGTCTGATTTAGGAATGTAATGAT 163
QY 319 TTTGTAATCTTAACAACAGATGCAATAGTTTAAATAAATCTTATCTAACTTTTAAAAA 378
Db 164 TCTGTATTAAATGTTAAATAAGATTATCTATTCGAAAAAGATATTTTCAACCTAAAAA 223
QY 379 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406
Db 224 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 251
```

RESULT 4  
US-09-764-846-30  
; Sequence 30, Application US/09764846  
; Patent No. US20020102638A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT212  
; CURRENT APPLICATION NUMBER: US/09/764,846  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 325  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-846-30

```
Query Match 13.6%; Score 55.2; DB 10; Length 325;
Best Local Similarity 60.8%; Pred. No. 0.016;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 259 CCGTTTTTATTGAATATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCATCTGAATAT 318
Db 104 CCTTTATAATTCACATAAATAAAGCATCTATTAGTGTCTGATTTAGGAATGTAATGAT 163
QY 319 TTTGTAATCTTAACAACAGATGCAATAGTTTAAATAAATCTTATCTAACTTTTAAAAA 378
Db 164 TCTGTATTAAATGTTAAATAAGATTATCTATTCGAAAAAGATATTTTCAACCTAAAAA 223
QY 379 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 406
Db 224 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 251
```

RESULT 5  
US-10-091-483-105  
; Sequence 105, Application US/10091483

```
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212C1
; CURRENT APPLICATION NUMBER: US/10/091.483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (612)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (620)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (621)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-091-483-105
```

```
Query Match      13.6%; Score 55.2; DB 9; Length 621;
Best Local Similarity 60.8%; Pred. No. 0.02;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 259 CCGTTTATTCGAATATCGTCAGTTTCCAAATTAAGTCAATTCGAGCCACTACTGAATAAT 318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 388 CCTTTAATTCACATAAAATAAGCATCTATTAGTCTGATTTAGGAATGTAAATGAT 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 TTGTGAATCTAACACAGATGCAATAGTTTAAATAAACTTATACCTTTTAAAAAAA 378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 448 TCTGTATTAAATGTAATAAGATTATCTATTGCAAAAAGATATTTCAACCTTAAAAAAA 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 535
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 6
US-09-764-846-105
; Sequence 105, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764.846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (612)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (620)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: SITE
; LOCATION: (621)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-105

Query Match      13.6%; Score 55.2; DB 10; Length 621;
Best Local Similarity 60.8%; Pred. No. 0.02;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 259 CCGTTTATTCGAATATCGTCAGTTTCCAAATTAAGTCAATTCGAGCCACTACTGAATAAT 318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 388 CCTTTAATTCACATAAAATAAGCATCTATTAGTCTGATTTAGGAATGTAAATGAT 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 TTGTGAATCTAACACAGATGCAATAGTTTAAATAAACTTATACCTTTTAAAAAAA 378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 448 TCTGTATTAAATGTAATAAGATTATCTATTGCAAAAAGATATTTCAACCTTAAAAAAA 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 AAAAAAATAAAAAAAAAAAAAAAAAAAAA 535
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 7
US-09-829-481-7
; Sequence 7, Application US/09829481
; Patent No. US20020069427A1
; GENERAL INFORMATION:
; APPLICANT: Presnail, James
; APPLICANT: Weng, Jude
; TITLE OF INVENTION: Arthropod Defensins
; FILE REFERENCE: BR1441 US NA
; CURRENT APPLICATION NUMBER: US/09/829.481
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/197279
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Argiope sp.
US-09-829-481-7
```

```
Query Match      13.4%; Score 54.6; DB 10; Length 351;
Best Local Similarity 54.9%; Pred. No. 0.021;
Matches 130; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

QY 23 TTCTCCACTCAGCTTCAAGAATGAAATCCATAGCTATTATTTTCATCGTTCTGTGGCC 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7 TCTGTCGACATTTCCAAAAAATCAATCGAGAGCTTCTGTTGTGATCTGCCTAGT--C 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 TTCTGTATTTTGGAGGATGGGATTTAGAACCTGGTTTGGATGTCCTTTTATGCGAGGA 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 GTCTGTGCTTTTGGCCAGAGTGACCGTGGAGCTGGTTTCGGCTGCCCTTCGACAGATG 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 AATGCCATAGACATTCGAAAAGTATTCGTCGTAGAGAGGCTTTTCGACAGGAACCTTC 202
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124 CAGTGCACATCATTCGAGGAGCATCAATATAGGGAGGAGTACTGCACCACTTATTC 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 AGGACAACTCGCTTTGCTATAGGTGAAATCCGATTTATTTGGCATAATGGAGACC 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 AAGCGACCTGCAAGTGTACGGATGATGACCCCTCCCTCTCTCAGACAGGAGCCCC 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
US-09-834-975-451/c
; Sequence 451, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
```



; SEQ ID NO 14052

; LENGTH: 471

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(471)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-14052

Query Match

Best Local Similarity 13.2%; Score 53.6; DB 9; Length 471;

Mismatches 89; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 257 ACCGGTTTTTAAATGATCGTCAGTTTCCAAATTAAGTCATTCGAGCCATCTGAATA 316

Db 296 ACACAATTTTGAAGAAACTGTCATCAGCTTATAACGACAATGTGCACTTTATAAATA 355

Qy 317 ATTTTGTAATCTAACACAGATGCAATAGTTTAAATAAACTTATACCTTTTAAAAA 376

Db 356 CTTCGACAGACTTTTANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 415

Qy 377 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406

Db 416 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 445

RESULT 12

US-09-960-352-12183/c

; Sequence 12183, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 12183

; LENGTH: 236

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 52-LIB3058-016-Q1-K1-E12

US-09-960-352-12183

Query Match

Best Local Similarity 13.1%; Score 53.2; DB 10; Length 236;

Mismatches 97; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 237 ATTATTTCGCATAATGGAGACCGGTTTTTATTGAATATGCTGAGTTTCCAAATTAAGTC 296

Db 220 AATTTTTCGGGTATATTTTCCAGTACATTTTGTGATTTTCCACTAACATTTTATTT 161

Qy 297 ATTCGAGCCATCTACTGAATAATTTTGTAACTAACACAGATGCAATAGTTTAAATAAAC 356

Db 160 TTTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 101

Qy 357 TTATACTTAATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406

Db 100 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 51

RESULT 13

US-09-960-352-13784/c

; Sequence 13784, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 13784

; LENGTH: 397

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 59-LIB3057-006-Q1-K1-G4

US-09-960-352-13784

Query Match

Best Local Similarity 13.1%; Score 53; DB 10; Length 397;

Mismatches 104; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 218 TGCATATAGTGAAATCCGATTTATTTGCCATAATGGAGACCGGTTTTTATTGAATATCG 277

Db 252 TTCAGTATATTAATAATTTATTTTATCATTATATAAAAAAAAAATTTTTTATAAACAAAA 193

Qy 278 TCAGTTTCCAATTAAGTCATTTTCGAGCCATCTGAATTAATTTTGTAACTTAACACAGA 337

Db 192 AATTTTATTAATATATGAATCTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 133

Qy 338 TGCATATAGTTTAAATAAATCTATATCTTAACTTTTAAAAAATAAAAAAAAAAAAAA 397

Db 132 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 73

Qy 398 AAAAAAAAAA 406

Db 72 AAAAAAAAAA 64

RESULT 14

US-09-960-352-5187

; Sequence 5187, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 5187

; LENGTH: 393

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 22-LIB34-074-Q1-E1-F5

US-09-960-352-5187

Query Match

Best Local Similarity 13.0%; Score 52.8; DB 10; Length 393;

Mismatches 81; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 279 CAGTTTCCAATTAAGTCATTTTCGAGCCATCTGAATTAATTTTGTAACTTAACACAGAT 338

Db 259 CAGGTTCCAATAAAAAACCAATTTGTAGACTCTGAAAAAATAAAAAAAAAAAAAA 318

Qy 339 GCAATAGTTTAAATAAATCTTACTTAACTTTTAAAAAATAAAAAAAAAAAAAA 398

Db 319 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 378

Qy 399 AAAAAAAA 406

Db 379 AAAAAAAA 386





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 18:05:58 ; Search time 1569 seconds  
(without alignments)  
4190.802 Million cell updates/sec

Title: US-09-829-481-3

Perfect score: 406

Sequence:

1 ctcctactacaatcactaagt.....aaaaaaaaaaaaaaaaaaaaa 406

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.4	15.4	355	BI323340	BI323340 kt69h08.y
2	62.2	15.3	328	BM154538	BM154538 fv86c10.y
3	62.2	15.3	329	AL513719	AL513719 AL513719
4	61.8	15.2	376	BM522187	BM522187 ESS00725
5	61.2	15.1	759	CNS060XV	AL411257 T7 end of
6	60.8	15.0	181	AI529496	AI529496 va52d02.y

7	60.2	14.8	470	13	BG928759
8	60.2	14.8	605	13	BG928759
9	60.2	14.8	1101	13	BG928759
10	59.8	14.7	284	13	BI745513
11	59.8	14.7	503	9	AL513809
12	59.2	14.6	186	14	C84834
13	59	14.5	326	13	BM154844
14	59	14.5	329	9	AL513719
15	59	14.5	424	13	BM263207
16	58.8	14.5	455	13	BI073560
17	58.2	14.3	265	14	C91235
18	58.2	14.3	631	10	AV729934
19	58	14.3	1182	14	BM906534
20	57.8	14.2	427	14	BQ638293
21	57.8	14.2	787	17	CNS049YD
22	57.8	14.2	1043	17	CNS0145P
23	57.8	14.2	1101	17	CNS017V2
24	57.6	14.2	256	14	BM966539
25	57.4	14.1	348	13	BM186008
26	57.4	14.1	366	10	AW101676
27	57.4	14.1	425	9	AL514791
28	57.4	14.1	943	12	BG392408
29	57.2	14.1	191	14	BQ637334
30	57.2	14.1	453	10	BE578190
31	57.2	14.1	866	13	BM015504
32	57.2	14.1	887	12	BG166263
33	57.2	14.1	924	17	CNS07A5L
34	57	14.0	438	9	AL634908
35	56.8	14.0	205	10	BE666934
36	56.8	14.0	481	9	AU037107
37	56.6	13.9	633	10	AW645696
38	56.4	13.9	166	14	BQ421843
39	56.4	13.9	209	14	BM882890
40	56.4	13.9	298	13	BI703836
41	56.4	13.9	352	13	BM154632
42	56.4	13.9	372	13	BM089774
43	56.4	13.9	626	9	AL513937
44	56.4	13.9	784	9	AL048397
45	56.2	13.8	349	9	AU033553

ALIGNMENTS

RESULT 1	BI323340	355 bp	mrna	linear	EST 30-JUL-2001				
LOCUS	BI323340	355 bp	mrna	linear	EST 30-JUL-2001				
DEFINITION	kt69h08.y1 Strongyloides ratti L2 pAMPI v1 Chiapelli McCarter								
	Strongyloides ratti cDNA 5' similar to TR:Q18611 Q18611 CODED FOR								
	BY C. ELEGANS CDNA CESEK63F. [1] ;contains element MER30 repetitive								
	element ;, mRNA sequence.								
ACCESSION	BI323340	GI:15002526							
VERSION	BI323340	EST							
KEYWORDS	Strongyloides ratti								
SOURCE	Strongyloides ratti								
ORGANISM	Strongyloides ratti								
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;								
AUTHORS	Panagrolaimoidea; Strongyloidea; Strongyloidea; Strongyloidea;								
	1 (bases 1 to 355)								
	McCarte, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,								
	Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,								
	Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, R.,								
	Runko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe								
	, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.								
	, Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and								
	Wilson, R.								
TITLE	The Washington Univ. Nematode EST Project, 1999								
JOURNAL	Unpublished (1999)								
COMMENT	Contact: McCarter JP								
	The Washington Univ. Nematode EST Project, 1999								
	Washington University School of Medicine								
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA								
	Tel: 314 286 1800								



## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 329)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. .329

## FEATURES

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CL0BA0072H01"

/clone\_lib="LTI\_NFL006\_PL2"

/tissue\_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 186 a 13 c 3 g 60 t 67 others

## ORIGIN

Query Match 15.3%; Score 62.2; DB 9; Length 329;  
 Best Local Similarity 44.1%; Pred. No. 8.7;  
 Matches 64; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 262 TTTTATGATATCGTCAGTTTCCAAATTAAAGTCATCTTCGAGCCATCTGAATAATTTT 321

Db 171 TTTTATGATATCGTCAGTTTCCAAATTAAAGTCATCTTCGAGCCATCTGAATAATTTT 112

QY 322 GTAATCAACACAGATGATAGTTTAAATAAATATATATCTTTTAAATAAATAAATAA 381

Db 111 WTWTWTWTAAWAWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 52

QY 382 AAAAAAAAAAAAAAAAAAAAAA 406

Db 51 AAAAAAAAAAAAAAAAAAAAAA 27

## RESULT 4

## BM522187

## LOCUS

BM522187 ESS00725 S.scabiei cDNA library Sarcptes scabiei cDNA clone

SAS0859 5', mRNA sequence. 376 bp mRNA linear EST 19-FEB-2002

Accession BM522187

Version BM522187.1 GI:18706427

Keywords EST.

Source Sarcptes scabiei.

Organism Sarcptes scabiei

Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

Acariformes; Sarcptiformes; Astigmata; Sarcptioidea; Sarcptidae;

Sarcptes.

1 (bases 1 to 376)

Ljunggren, E.L., Nilsson, D., Naslund, K. and Mattsson, J.G.

Expressed sequence tag analysis of the parasitic mite Sarcptes

scabiei

Unpublished (2001)

Contact: Mattsson J.G.

Department of Parasitology (SNEPAR)

National Veterinary Institute

SE-751 89 Uppsala, Sweden

Tel: +46 18 674120

Fax: +46 18

Email: jens.mattsson@sva.se

Similar to gi|8777580|dbj|BAA97098.1| (AP002460) gene\_id:F1D9.26

unknown protein [Arabidopsis thaliana]

Seq primer: T3 primer

High quality sequence stop: 376.

## FEATURES

source

Location/Qualifiers

1. .376

/organism="Sarcptes scabiei"

/db\_xref="taxon:52283"

/clone="SAS0859"

/note="The Sarcptes scabiei mixed life stage library was

constructed by Jens G Mattsson. cDNAs were synthesized and

from poly(A)+ RNA by oligo d(T) priming, size-selected and

directionally cloned into the Uni-ZAP lambda vector

(Stratagene). The primary library was amplified on

XL1-Blue MRF' cells."

BASE COUNT 165 a 48 c 60 g 97 t 6 others

## ORIGIN

Query Match 15.2%; Score 61.8; DB 13; Length 376;  
 Best Local Similarity 57.5%; Pred. No. 9;  
 Matches 111; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 214 CGTTTGCTATAGTGAAATCCGATTTATTCATATGAGACCCGTTTATTGAT 273

Db 141 CGATTGCGAATTATAAATCTGATTAAATTTAACTCTAAAGACACTGATTCAC 200

QY 274 ATCGTCAGTTTCCAAATTAAGTCATTTTCGAGCCATCTGATTAATTTGTAATCTAACAA 333

Db 201 TTTTTCATTTTCTTCAATCGGGTAAATTTCTCGAACTATTCGAAATAAAAAA 260

QY 334 CAGATGCAATAGTTTAAATAAATCTATATCTTAACTTTTAAAAAATAAAAAA 393

Db 261 GTATTTAAAGATTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 320

QY 394 AAAAAAAAAAAAAA 406

Db 321 AAAAAAAAAAAAAA 333

## RESULT 5

## CNS060XV

## LOCUS

CNS060XV T7 end of clone AW0AA009H09 of library AW0AA from strain CLIB 89 of

Yarrowia lipolytica, genomic survey sequence. 759 bp DNA linear GSS 05-JUL-2001

Accession AL411257

Version AL411257.1 GI:12180512

Keywords GSS.

Source Yarrowia lipolytica.

Organism Yarrowia lipolytica

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Dipodascaceae; Yarrowia.

1 (bases 1 to 759)

Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durren, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neveu, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 759)

Casaregola, S., Neveu, C., Lepingle, A., Bon, E., Feynrol, C.,

Artiguenave, F., Wincker, P. and Gaillardin, C.

Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia

lipolytica

JOURNAL MEDLINE 20584727

PUBMED 11152892

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
3 (bases 1 to 759)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segrer@genoscope.cns.fr - web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
source
1..759
/organism="Yarrowia lipolytica"
/strain="CLIB 89"
/db_xref="taxon:4952"
/clone="AW0AA009H09"
/clone_lib="AW0AA"
/note="end : T7"
BASE COUNT
356 a 53 c 51 g 206 t 93 others
ORIGIN
Query Match 15.1%; Score 61.2; DB 17; Length 759;
Best Local Similarity 49.4%; Pred. No. 6.8;
Matches 84; Conservative 24; Mismatches 62; Indels 0; Gaps 0;
Qy 237 ATTATATTCGCAATATGAGACCGCTTTTATTGAATATCGTCAGTTCCCAATTAAGTC 296
||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 155 ATWTATATWTATATATATATATATATATATATATATATATATATATATATATAT 214
Qy 297 ATTTGCGCCATACGTAATATTTTGTATCTAACACAGATGCAATAGTTTAAATAAAC 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 ATWTATATATATATATATATATATATATATATATATATATATATATATATATAT 274
Qy 357 TTATCTACTTACTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 324

RESULT 6
AI529496 181 bp mRNA linear EST 18-MAR-1999
LOCUS va52d02.v1 Soares mouse 3NME12 5 Mus musculus cDNA clone
DEFINITION IMAGE:734979 5', mRNA sequence.
ACCESSION AI529496
VERSION AI529496.1 GI:4443631
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 181)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:452027

```

```

Seq primer: -40RP from Gibco
High quality sequence stop: 91.
Location/Qualifiers
source
1..181
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:734979"
/clone_lib="Soares mouse 3NME12 5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/dev_host="DH10B"
/note="Organ: whole fetus; Vector: p7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTCAAGTGGGAGCGCGCTTATTTTATTTTATTTTATTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaudo. "
BASE COUNT
110 a 19 c 18 g 34 t
ORIGIN
Query Match 15.0%; Score 60.8; DB 9; Length 181;
Best Local Similarity 80.7%; Pred. No. 20;
Matches 71; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Qy 319 TTGTGAATCTAACACACATGATGTTTAAATAAATCTTACTTAACTTTTAAAAA 378
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 TTGTGAATCTTACAGAGATGTAATTTTGAATATAAATCTTAACTTAACTTAA 113
Qy 379 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 406
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 141

RESULT 7
BG928759 470 bp mRNA linear EST 06-NOV-2001
LOCUS HNC56-1-C12.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
DEFINITION BG928759.1 GI:14323282
ACCESSION BG928759
VERSION BG928759.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 470)
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries
Osteoarthritis. Cartil. 9 (7), 641-653 (2001)
21482651
Contact: Sanjay Kumar
UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-legsk.com
Seq primer: T7.
Location/Qualifiers
source
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"

```

QY	322	GTATCTTAACAACAGATGCATAGTCTTTAAATAAAGCTTATACTTAACTTTTAAAAA	381
Db	265	GCCTGCCAAAAA	324
QY	382	AAAAAAAAAAAAAAAAAAAAA	406
Db	325	AAAAAAAAAAAAAAAAAAAAA	349
RESULT 9			
CNS0021J			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	230	AAATCCGATTATTGGCATATATGAGACCCGTTTTTTATTGAATATCGTCGATTTCCAAAT	289
Db	666	AWAAATTTCTTTATWATWAAATWAAATTTTAAAAAATTTTTTTTAWTTTTTTTTTAAAAA	725
QY	290	TAAATGCTATTCGACCCATCTAGTAATTTTGTAATCTAACACAGATGCAATAGTTTA	349
Db	726	TTTATTATWTTTTAAANATWATWTTTAAATTTTTTATWATATWATWATWATWATWATWAT	785
QY	350	AATAAATCTTACTACTTTTAAAAA	406
Db	786	AATTTAAATTTTTTTTAAATAAAAA	842
RESULT 10			
B1745513			



**COMMENT**

Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
Tel: 81-298-53-4664  
Fax: 81-298-53-6614  
Email: hideko@biol.tsukuba.ac.jp  
PROJECT = "dictyostelium discoideum cdna" project

[illegible]

```
location/Qualifiers
1 186
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:46889"
/clone="SSG726"
/clone_1lb="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slus"
```

Query Match 14.6%; Score 59.2; DB 14; Length 186;  
Best Local Similarity 76.0%; Pred. No. 33;  
Matches 73; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

RESULT 13  
BM154844

LOCUS	BM154844	326 bp	mRNA	linear	EST 03-DEC-2001
DEFINITION	fv90c01.y1 zebrafish SJD adult male II Danio rerio cDNA clone 5544193.5, mRNA sequence.				

**SOURCE ORGANISM**

zebrafish.  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.  
1 (bases 1 to 326)  
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy  
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood  
, K., Steptoe, M., Thelshing, B., Allen, M., Bowers, Y., Person, B.,  
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.  
and Wilson, R.

TITLE	COMMENT
Washo Zebrafish EST Project 1998 and Wilson, K.	Unpublished (1998) Contact: Stephen L. Johnson Washington University School of 4444 Forest Park Parkway, Box 8510 Tel: 314 286 1800 Fax: 314 286 1810

FEATURES  
SOUND

```

1. .326
/organism="Danio rerio"
/strain="SJD"
/db_xref="taxon:7955"
/clone="5544193"

```

```

/clone_lib="zebrafish SJD adult male II"
/sex="male"
/tissue_type="whole body"
/lab_host="DH10B"
/notes="Vector: pAMP1; Site_1: EcoRI; Site_2: NotI; First
strand cDNA synthesis was primed using oligo-dT on
magnetic beads with an additional primer .
5'-ggcgccgaataacgactacta-taggg-3'. Second strand
synthesis was a 3-cycle PCR using the primers
5'-ggcgctaatcagctactactag-3' and
5'-aacgagtgttaacacgacagactctt-tttttttttn-3'. The
cDNA was subsequently amplified in a 7-cycle PCR with the
following primers: 5'-ggcgctactactactactag-3' and
5'-aacgagtgtt-aacacgacag. Deoxy-UMP adaptors were added in
a third PCR (5 cycles) and the primers
5'-cauacacaucaugccgcgtactacgactactaggg-3' and
5'-cauacacuaaagcagtggttaacacgacagatgc-3'. Ends were
treated with uracil DNA glycosylase and product with 3'
overhangs was annealed to complementary ends of pAMP1.
Insert can be excised using EcoRI and NotI. Library
constructed by Joe Barnes and Steve Johnson (Washington
University)."

```

Query Match	14.5%	Score 59;	DB 13;	Length 326;
Best Local Similarity	52.7%	Pred. No. 24;		

Qy	224	AGTGAAATCCGATTATTATTTGCCATAATGGAGACCGTTTTATTTGAATATCGTCAGTT	283
Db	135	TGTTAATAATCATTTTACTGTGTACAAAAGGCAGACCACCTGTCTAATGAAAAGTCAATA	194
Qy	284	TCCAATTAAAGTCATTTTCGAGCCATACTGNAATATTTTGTAACTCTAACACAGATGCAAT	343
Db	195	AACATGAAAGACATTATGATGGAAAAAATAAAAAAGAAAAAATAAAAAAATAAAAAA	254
Qy	344	AGTTTAAATAAATCTTACTTTAACTTTTAAAAAATAAAAAAATAAAAAAATAAAAAA	403
Db	255	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA	314
Qy	404	AAA 406	
Db	315	AAA 317	

RESULT	14
AL513719	
LOCUS	
DEFINITION	AL513719 LTI_NFL006_PL2 329 bp mRNA linear EST 13-FEB-2001
	prime, rRNA sequence. Homo sapiens cDNA clone CL0BA0072H01.3
ACCESSION	AL513719
VERSION	AL513719.1
KEYWORDS	GI:12777213
SOURCE	EST.
	human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 329)
TITLE	Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
JOURNAL	Full-length cDNA libraries and normalization Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage Bp 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES	Location/Qualifiers 1..329 /organism="Homo sapiens"
source	

